电池数据分析平台\_使用说明

 上海翌擎

## 背景介绍

尽快实现相关电池预警模型的工程化实现，以便应用到目前新能源大批量的车辆数据上，同时需要在完成监控平台车辆运行指标参数计算的基础上，开发电池故障之间的关联关系模型。以此做到预警报警双管齐下，增强车辆热失控领域的技术能力；关注电池健康度（SOH）,能够及时提供不同时间窗口及工况下，电池的健康状况，提高客户满意度，减少不必要的“趴窝”现象。最终确保 新能源的车辆安全行驶，降低维护成本、提高产品销量，并且为守护用户生命安全、提升产品品牌形象并且满足国家监管要求提供数据支撑。

## 文档编写目标

本次文档编写的目标主要有以下几个方面：

1.信息传递：确保文档能够清晰、准确地传递必要的信息给目标读者。

2.指导操作：提供详细的模型步骤和指南，帮助用户完成特定的任务或操作。

3.技术支持： 提供技术支持信息，帮助用户解决技术问题。

## 售后支持

电池算法模型产品的售后支持条款是确保用户在使用算法模型产品过程中能够获得及时、有效的技术支持和服务保障的重要文件。以下是电池算法模型产品售后支持条款

#### 一、服务范围

1. **技术支持**：提供电池算法模型产品的技术咨询、问题解答、故障排除等技术支持服务。
2. **培训服务**：为用户提供电池算法模型产品的使用培训、操作指导等，确保用户能够熟练使用产品。
3. **升级与维护**：根据产品更新情况，为用户提供电池算法模型产品的升级服务，以及必要的维护服务。

#### 二、服务期限

1. **服务有效期**：本售后支持条款自用户购买电池算法模型产品之日起生效，有效期为**产品购买后的一年时间。**
2. **续签与延期**：服务期限届满后，用户可根据需要选择续签或延期服务，具体费用及条款以双方协商为准。

#### 三、服务响应与处理

1. **服务请求提交**：用户可通过ai service@nx-engine.com方式向售后服务团队提交服务请求。
2. **响应时间**：售后服务团队在接到用户服务请求后，应在约定时间内24小时内给予响应，并根据情况提供相应的解决方案或处理建议。
3. **处理流程**：售后服务团队将按照约定的处理流程，对用户提出的问题或需求进行逐一处理，确保问题得到及时解决。

#### 四、备件与替换

1. **备件提供**：在电池算法模型产品出现故障且无法修复时，售后服务团队将根据实际情况为用户提供备用模型替换服务。
2. **替换条件**：备用模型替换需符合产品保修条款及双方协商的替换条件。

#### 五、保修与免责

1. **保修范围**：电池算法模型产品在正常使用情况下出现的故障，且非因用户不当使用或第三方因素导致的损坏，属于保修范围。
2. **免责条款**：因用户不当使用、第三方因素或不可抗力导致的算法模型产品损坏，不在保修范围内。同时，售后服务团队不承担因用户未按照产品说明或操作指南使用产品而导致的任何损失或责任。

#### 六、争议解决与适用法律

1. **争议解决**：双方因执行本售后支持条款产生的争议，应首先通过友好协商解决；协商不成的，可提交至相关仲裁机构或法院解决。
2. **适用法律**：本售后支持条款的解释、效力及解决争议的适用法律均为相关法律法规。

#### 七、其他条款

1. **条款变更**：售后服务团队有权根据业务发展需要或法律法规要求变更本售后支持条款，变更后将及时通知用户。
2. **用户同意**：用户在使用算法模型产品时，即视为同意本售后支持条款的所有内容。
3. **保密义务**：双方应对在合作过程中获取的对方商业秘密及未公开信息承担保密义务，未经对方书面同意，不得向第三方披露或用于其他用途。

# 电池数据分析模型运行架构图

## 大数据平台技术



### 产品部署流程图



### 产品安装部署说明

1. 导入依赖包以及spark初始化

注：导入python依赖包，以及使用spark的版本、spark的python环境依赖、spark的一些参数设置等

# 导入必要的依赖包

# 设置SparkSession的配置

1. 外部资源初始化和导入

注：初始化spark、redis；导入算法模型

# 初始化SparkSessio

# 导入算法模型

# 初始化redis

1. 导入批数据的数据源

注：导入所需的原始数据（离线的批数据，一般是最小是几天，最多可能是整个生命周期的数据）

1. 模型推理

注：模型推理，包含数据预处理、特征提取、模型预测等功能，最终是从原始数据到结果数据的整个流程

 【获取充电片段数据】

 :param x: 单车充电数据 (Dataframe)

 :param start\_soc: 充电开始最高SOC限制 (int)

 :param end\_soc: 充电结束最低SOC限制 (int)

 :return: 单车的多次充电片段数据（list）

 【获取模型相关特征】

 :param charge\_data: 电池充电片段数据（list）（来源于get\_charge函数的返回值）

 :param start\_date: 起始充电时间（str）

 :param circle\_ac: 累计循环次数（float）

 :param ars\_use\_behavior\_score\_ac: 累计使用行为得分（int）

 :return: soh相关特征（DataFrame），包括vin码、时间戳、累计里程、日历天数、循环次数、车辆使用行为

 # 去除异常累计里程的影响（）同时也能去除其他因素的异常值影响

 【当前模型预测】

 :param soh\_feature: soh的特征数据（DataFrame)

 :return: soh结果（DataFrame）包含三列，分别是vin码，时间，SOH

1. 调用语句

注：原始数据进行功能函数的调用（分布式），进而得到算法结果

1. 结果入库

注：将最终的结果数据写入到指定的存储介质中

### 环境说明

| **资源名称** | **版本号** | **用途及说明** |
| --- | --- | --- |
| HDFS | 2.6.0 | 分布式文件系统 |
| Hive | 1.1.0 | 数据仓库工具 |
| Kafka | 0.9.0 | 消息队列服务 |
| Task Scheduler | 5.2.1 | 任务调度-联想自研组件 |
| MapReduce | 2.6.0 | 批量分布式计算 |
| Spark | 2.0.0 | 内存与实时分析框架 |
| YARN | 2.6.0 | 资源管理和调度 |
| ZooKeeper | 3.4.5 | 分布式应用程序协调服务 |
| JDK | 1.8 | JDK |
| Python | 2.6.6/3.5.9 | Python |
| MYSQL | 5.7 | 关系数据库 |
| FineBI | 4.1/5.1 | 多维数据分析可视化工具 |

### 平台组件

 新能源大数据平台电池分析系统的大数据平台依赖于大数据平台进行开发。

安装组件如下：

1. HDFS

Hadoop分布式文件系统(HDFS)是指被设计成适合运行在通用硬件(commodity hardware)上的分布式文件系统（Distributed File System）。它和现有的分布式文件系统有很多共同点。但同时，它和其他的分布式文件系统的区别也是很明显的。HDFS是一个高度容错性的系统，适合部署在廉价的机器上。HDFS能提供高吞吐量的数据访问，非常适合大规模数据集上的应用。

1. YARN

Apache Hadoop YARN （Yet Another Resource Negotiator，另一种资源协调者）是一种新的 Hadoop 资源管理器，它是一个通用资源管理系统，可为上层应用提供统一的资源管理和调度，它的引入为集群在利用率、资源统一管理和数据共享等

1. Zookeeper

ZooKeeper是一个[分布式](https://baike.baidu.com/item/%E5%88%86%E5%B8%83%E5%BC%8F/19276232)的，开放源码的[分布式应用程序](https://baike.baidu.com/item/%E5%88%86%E5%B8%83%E5%BC%8F%E5%BA%94%E7%94%A8%E7%A8%8B%E5%BA%8F/9854429)协调服务，是[Google](https://baike.baidu.com/item/Google)的Chubby一个[开源](https://baike.baidu.com/item/%E5%BC%80%E6%BA%90/246339)的实现，是Hadoop和HBase的重要组件。它是一个为分布式应用提供一致性服务的软件，提供的功能包括：配置维护、域名服务、分布式同步、组服务等。

1. Spark

Apache Spark 是专为大规模数据处理而设计的快速通用的计算引擎。Spark是UC Berkeley AMP lab (加州大学伯克利分校的AMP实验室)所开源的类Hadoop MapReduce的通用并行框架，Spark，拥有Hadoop MapReduce所具有的优点；但不同于MapReduce的是——Job中间输出结果可以保存在内存中，从而不再需要读写HDFS

1. Hive

Hive是一个数据仓库基础工具在Hadoop中用来处理结构化数据。它架构在Hadoop之上，总归为大数据，并使得查询和分析方便。并提供简单的sql查询功能，可以将sql语句转换为MapReduce任务进行运行。

1. Kafka

Apache Kafka 是一个开源分布式事件流平台，被数千家公司用于高性能数据管道、流分析、数据集成和关键任务应用程序。

### 任务管理

1. 任务调度平台

 新能源大数据平台电池分析系统的调度平台使用的是 内部联想自研组件调度平台 。

### 数据生产

 新能源大数据平台电池分析系统车辆数据由车载TBOX按照周期自动上传到送到kafka集群中。 大数据监控系统收集 新能源大数据平台电池分析系统车辆的数据分析的最终呈现结果为：帆软BI展示相关的信号分析结果做可视化web页面展示。

### 数据计算

1. 离线计算

采用SparkCore或是SparkSQL 进行离线计算将结果输入到数据仓库中，离线计算平台任务可以按照天、周、月以及自定义时间进行一次或多次任务执行，并可以重复执行，以及补充历史数据等。

1. 数据架构

数据仓库模型如下：

* ODS（操作数据）

或是原始数据层，用于存储 数据，或是其他来源数据

* DWS（公共明细）

对ODS层数据进行清洗（去除空值，脏数据，超过极限范围的数据）

* DWD（公共汇总）

可以对应数据进行维度汇总形成各种组合表

* ADS（数据应用）

这个是数据仓库的最后一层数据，为应用层数据，直接可以给业务人员使用

### 数据存储

1. Clickhouse
* 新能源大数据平台电池分析系统最终计算的结果数据存储到clickhouse中（近六个月的数据）
* 指标数据
1. MySQL
* 规则配置数据
1. Hive
* 数据仓库的存储
* 可以导出原始数据
* 可以通过 内部任务调度平台界面执行SQL查询展示数据数据

### 数据可视化

 新能源大数据平台电池分析系统数据可视化采用帆软BI, finebi是B/S架构，在浏览器端对数据进行分析和查看，以及系统平台的管理