



(12) 发明专利

(10) 授权公告号 CN 107402976 B

(45) 授权公告日 2020.10.30

(21) 申请号 201710530630.3

(22) 申请日 2017.07.03

(65) 同一申请的已公布的文献号
申请公布号 CN 107402976 A

(43) 申请公布日 2017.11.28

(73) 专利权人 国网山东省电力公司经济技术研究院

地址 250021 山东省济南市槐荫区纬十路111号

专利权人 国家电网公司

(72) 发明人 李雪亮 吴奎华 蒋斌 蒋忠福
吴健 安勇 李山 冯亮 杨波
贾善杰 汪有杰 刘波

(74) 专利代理机构 济南诚智商标专利事务有限公司 37105

代理人 李修杰

(51) Int.Cl.

G06F 16/182 (2019.01)

G06F 16/17 (2019.01)

G06F 16/27 (2019.01)

G06K 9/62 (2006.01)

G06Q 50/06 (2012.01)

(56) 对比文件

CN 103700032 A, 2014.04.02

CN 105787089 A, 2016.07.20

CN 105824847 A, 2016.08.03

CN 105278373 A, 2016.01.27

US 2003142851 A1, 2003.07.31

审查员 杨欢

权利要求书3页 说明书10页 附图3页

(54) 发明名称

一种基于多元异构模型的电网多源数据融合方法及系统

(57) 摘要

本发明公开了一种基于多元异构模型的电网多源数据融合方法及系统,方法包括以下过程:多源数据接入;元数据存储;设备统一建模及数据提取;模型匹配度计算;多源数据融合。系统主要包括多源数据接入、元数据存储、设备统一建模及数据提取、模型匹配度计算和多源数据融合五个模块。本发明通过深入分析多源系统数据,建立包括各源系统数据统一模型,通过模型遍历比对,计算模型间匹配度,根据匹配度情况进行自动关联,实现了90%以上的数据自动集成融合,并准确分析出未能融合数据明细及具体原因,既实现了各业务数据的集成融合,又对整体的融合质量进行了有效的管控。



1. 一种基于多元异构模型的电网多源数据融合方法,其特征是,包括以下过程:

多源数据接入:采用多种技术手段进行结构化数据、非结构化数据、海量历史/准实时数据和电网空间数据的接入;

元数据存储:对各类数据按照统一数据规范进行标准化格式存储;

设备统一建模及数据提取:确定并统一需要进行数据融合的设备类型、电压等级范围和业务领域信息,建立多源数据的统一模型,并进行数据提取和比对计算;

模型匹配度计算:根据多源数据的统一模型进行设备匹配度计算,通过统一模型中的属性信息对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异;

多源数据融合:对于A类数据直接建立关联关系;对于B类数据则开发数据融合对应工具,用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系;对应C类数据则以问题的形式反馈至源系统进行数据修补,修补完成后再根据修补增量实现自动关联;

所述模型匹配度计算的具体过程为:

通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦;

在确保根节点已对应的情况下,对各系统中的逻辑模型进行横向匹配,计算各节点包含下级节点数量,并做为相应节点的属性;

通过有向图的拓扑排序算法计算拓扑模型中节点的前驱及后续节点关系,首先在有向图中选取一个没有前驱的顶点,将其输出从有向图中删除该顶点,并且删除以该顶点为尾的所有有向图的边;重复以上的步骤,直到图中的所有顶点均输出或是图中的顶点均没有前驱为止,确立该节点的上下文在各源系统模型中是否一致;

在对设备聚焦的基础上,通过各系统中特征模型中的属性信息,对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异。

2. 根据权利要求1所述的一种基于多元异构模型的电网多源数据融合方法,其特征是,所述的多源数据包括实时数据、经常变化的结构化数据、基本不会变化的结构化数据和非结构化数据;所述的多源数据接入的具体过程为:

通过Kafka分布式消息队列将实时数据接入大数据平台;

通过ETL工具Kettle、增量捕获工具将经常变化的结构化数据写入大数据平台的分布式关系型数据库中;

通过Sqoop工具将基本不会变化的结构化数据写入大数据平台的分布式非关系型数据库和分布式文件系统中;

通过Flume工具将非结构化数据直接写入大数据平台的分布式文件系统。

3. 根据权利要求2所述的一种基于多元异构模型的电网多源数据融合方法,其特征是,在设备统一建模及数据提取过程中,所述的设备类型包括至少变电站、主变、线路、配变和开关,所述的电压等级范围为500kV~10kV,所述的业务领域包括台账信息、运行信息、空间信息和规划信息。

4. 根据权利要求2所述的一种基于多元异构模型的电网多源数据融合方法,其特征是,

所述统一模型包括逻辑模型、拓扑模型和特征模型，

所述逻辑模型为以高电压等级变电站为根节点，按设备类型分类为分支，10kV配变为叶子节点的树形数据结构，做为各业务系统电网设备的逻辑模型；

所述拓扑模型为以电气接线关系、地理接线关系拓扑图，以及电压等级由高到底、设备类型统一的拓扑模型有向图数据结构，做为各业务系统电网设备的拓扑模型；

所述特征模型为根据设备共有属性信息提取设备特征模型线性结构，对设备类型进行统一编码，每个设备类型建立独立的对象模型，以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类，以每类设备特有属性为子类的特征模型。

5. 一种基于多元异构模型的电网多源数据融合系统，其特征是，包括：

多源数据接入模块，用以采用多种技术手段进行结构化数据、非结构化数据、海量历史/准实时数据和电网空间数据的接入；

元数据存储模块，用以对各类数据按照统一数据规范进行标准化格式存储；

设备统一建模及数据提取模块，用以确定并统一需要进行数据融合的设备类型、电压等级范围和业务领域信息，建立统一模型，并进行数据提取和比对计算；

模型匹配度计算模块，用以根据多源数据的统一模型进行设备匹配度计算，通过统一模型中的属性信息对每类设备进行细化匹配，各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类，并记录设备特征属性的差异；

多源数据融合模块，用以对于A类数据直接建立关联关系；对于B类数据则开发数据融合对应工具，用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系；对应C类数据则以问题的形式反馈至源系统进行数据修补，修补完成后再根据修补增量实现自动关联；

所述模型匹配度计算模块包括：

逻辑模型计算模块，用以通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦，以及在确保根节点已对应的情况下，对各系统中的逻辑模型进行横向匹配，计算各节点包含下级节点数量，并做为相应节点的属性；

拓扑模型计算模块，用以通过有向图的拓扑排序算法计算拓扑模型中节点的前驱及后续节点关系，首先在有向图中选取一个没有前驱的顶点，将其输出从有向图中删除该顶点，并且删除以该顶点为尾的所有有向图的边；重复以上的步骤，直到图中的所有顶点均输出或是图中的顶点均没有前驱为止，确立该节点的上下文在各源系统模型中是否一致；

特征模型计算模块，用以在对设备聚焦的基础上，通过各系统中特征模型中的属性信息，对每类设备进行细化匹配，各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类，并记录设备特征属性的差异。

6. 根据权利要求5所述的一种基于多元异构模型的电网多源数据融合系统，其特征是，所述的多源数据包括实时数据、经常变化的结构化数据、基本不会变化的结构化数据和非结构化数据；

所述的多源数据接入模块包括：

实时数据接入模块，用以通过Kafka分布式消息队列将实时数据接入大数据平台；

经常变化的结构化数据写入模块，用以通过ETL工具Kettle、增量捕获工具将经常变化

的结构化数据写入大数据平台的分布式关系型数据库中；

基本不会变化的结构化数据写入模块，用以通过Sqoop工具将基本不会变化的结构化数据写入大数据平台的分布式非关系型数据库和分布式文件系统中；

非结构化数据写入模块，用以通过Flume工具将非结构化数据直接写入大数据平台的分布式文件系统。

7. 根据权利要求6所述的一种基于多元异构模型的电网多源数据融合系统，其特征是，所述的设备类型包括至少变电站、主变、线路、配变和开关，所述的电压等级范围为500kV~10kV，所述的业务领域包括台账信息、运行信息、空间信息和规划信息。

8. 根据权利要求6所述的一种基于多元异构模型的电网多源数据融合系统，其特征是，所述统一模型包括逻辑模型、拓扑模型和特征模型，

所述逻辑模型为以高电压等级变电站为根节点，按设备类型分类为分支，10kV配变为叶子节点的树形数据结构，做为各业务系统电网设备的逻辑模型；

所述拓扑模型为以电气接线关系、地理接线关系拓扑图，以及电压等级由高到底、设备类型统一的拓扑模型有向图数据结构，做为各业务系统电网设备的拓扑模型；

所述特征模型为根据设备共有属性信息提取设备特征模型线性结构，对设备类型进行统一编码，每个设备类型建立独立的对象模型，以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类，以每类设备特有属性为子类的特征模型。

一种基于多元异构模型的电网多源数据融合方法及系统

技术领域

[0001] 本发明涉及一种基于多元异构模型的电网多源数据融合方法及系统,属于智能控制技术领域。

背景技术

[0002] 随着计算机技术在电力行业的不断发展,电力行业出现了大量专业化信息系统,贯穿发电、输电、变电、配电、用电各个环节,每个专业系统在各异的领域起到了举足轻重的作用,在国家电网公司系统内,PMS、EMS、GIS、智能配网监控平台,用电信息采集等系统分别在运检、调度、营销等领域得到了广泛应用,积累了大量数据资源,并分别通过数据中心以共享数据、共享服务等方式对外提供实时/准实时数据服务,提升数据资产的利用率。

[0003] 但是由于各专业信息系统所处的领域不同,关注的业务有较大差异性,原始数据来源不统一,数据存储结构多样化,导致同一个电网设备在不同的信息系统中编码、类型、名称、关联关系等数据结构存在较大差异,无法对同一个设备进行台账信息、运行信息、空间信息等数据全面的查询和统计,无法对区域电网情况形成多维度诊断、分析。

[0004] 电网规划业务具有数据源头多、数据体量大、统计维度多等特点,科学合理电网规划的基础是对各个层次、各个维度的海量数据信息进行集成融合、统筹管理,挖掘有效信息,以提升电网规划质量和效率。为支撑电网规划诊断分析、负荷分析预测、电网仿真计算、能源电力供需、经济技术评估等业务的开展,亟需对各系统共享的多元异构数据进行集成融合,实现了设备数据库、运行数据库、图形数据库、规划数据库的“四库合一”。

[0005] 受限于对海量数据的查询处理效率,传统的数据集成融合方式主要是依赖管理手段,协调各数据管理部门通过人工对应方式实现存量数据的关联融合,集成各数据管理部门业务流程实现增量数据的关联融合。这种方式虽然能够实现数据的融合,但是存在如下问题,一是电网设备体量大,数据对应工作繁琐,需耗费大量人力物力。二是各数据管理部门人员对其他部门数据模型、业务流程不熟悉,容易造成数据关联出错率高,关联效率低。三是各部门业务流程集成难度高,当各部门数据接口有变更升级时,数据融合总线要频繁更新升级,提高了个部门业务流程复杂度。

发明内容

[0006] 针对上述不足,本发明提供了一种基于多元异构模型的电网多源数据融合方法及系统,其不仅能够实现各业务数据的集成融合,而且能够对整体的融合质量进行了有效的管控。

[0007] 本发明解决其技术问题采取的技术方案是:

[0008] 一方面,提供了一种基于多元异构模型的电网多源数据融合方法,它包括以下过程:

[0009] 多源数据接入:采用多种技术手段进行结构化数据、非结构化数据、海量历史/准实时数据和电网空间数据的接入;

- [0010] 元数据存储:对各类数据按照统一数据规范进行标准化格式存储;
- [0011] 设备统一建模及数据提取:确定并统一需要进行数据融合的设备类型、电压等级范围和业务领域信息,建立多源数据的统一模型,并进行数据提取和比对计算;
- [0012] 模型匹配度计算:根据多源数据的统一模型进行设备匹配度计算,通过统一模型中的属性信息对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异;
- [0013] 多源数据融合:对于A类数据直接建立关联关系;对于B类数据则开发数据融合对应工具,用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系;对应C类数据则以问题的形式反馈至源系统进行数据修补,修补完成后再根据修补增量实现自动关联。
- [0014] 进一步地,所述的多源数据包括实时数据、经常变化的结构化数据、基本不会变化的结构化数据和非结构化数据;所述的多源数据接入的具体过程为:
- [0015] 通过Kafka分布式消息队列将实时数据接入大数据平台;
- [0016] 通过ETL工具Kettle、增量捕获工具将经常变化的结构化数据写入大数据平台的分布式关系型数据库中;
- [0017] 通过Sqoop工具将基本不会变化的结构化数据写入大数据平台的分布式非关系型数据库和分布式文件系统中;
- [0018] 通过Flume工具将非结构化数据直接写入大数据平台的分布式文件系统。
- [0019] 进一步地,在设备统一建模及数据提取过程中,所述的设备类型包括至少变电站、主变、线路、配变和开关,所述的电压等级范围为500kV~10kV,所述的业务领域包括台账信息、运行信息、空间信息和规划信息。
- [0020] 进一步地,所述统一模型包括逻辑模型、拓扑模型和特征模型,
- [0021] 所述逻辑模型为以高电压等级变电站为根节点,按设备类型分类为分支,10kV配变为叶子节点的树形数据结构,做为各业务系统电网设备的逻辑模型;
- [0022] 所述拓扑模型为以电气接线关系、地理接线关系拓扑图,以及电压等级由高到底、设备类型统一的拓扑模型有向图数据结构,做为各业务系统电网设备的拓扑模型;
- [0023] 所述特征模型为根据设备共有属性信息提取设备特征模型线性结构,对设备类型进行统一编码,每个设备类型建立独立的对象模型,以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类,以每类设备特有属性为子类的特征模型。
- [0024] 进一步地,所述模型匹配度计算的具体过程为:
- [0025] 通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦;
- [0026] 在确保根节点已对应的情况下,对各系统中的逻辑模型进行横向匹配,计算各节点包含下级节点数量,并做为相应节点的属性;
- [0027] 通过有向图的拓扑排序算法计算拓扑模型中节点的前驱及后续节点关系,首先在有向图中选取一个没有前驱的顶点,将其输出从有向图中删除该顶点,并且删除以该顶点为尾的所有有向图的边;重复以上的步骤,直到图中的所有顶点均输出或是图中的顶点均没有前驱为止,确立该节点的上下文在各源系统模型中是否一致;
- [0028] 在对设备聚焦的基础上,通过各系统中特征模型中的属性信息,对每类设备进行

细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异。

[0029] 另一方面,提供了一种基于多元异构模型的电网多源数据融合系统,它包括:

[0030] 多源数据接入模块,用以采用多种技术手段进行结构化数据、非结构化数据、海量历史/准实时数据和电网空间数据的接入;

[0031] 元数据存储模块,用以对各类数据按照统一数据规范进行标准化格式存储;

[0032] 设备统一建模及数据提取模块,用以确定并统一需要进行数据融合的设备类型、电压等级范围和业务领域信息,建立统一模型,并进行数据提取和比对计算;

[0033] 模型匹配度计算模块,用以根据多源数据的统一模型进行设备匹配度计算,通过统一模型中的属性信息对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异;

[0034] 多源数据融合模块,用以对于A类数据直接建立关联关系;对于B类数据则开发数据融合对应工具,用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系;对应C类数据则以问题的形式反馈至源系统进行数据修补,修补完成后再根据修补增量实现自动关联。

[0035] 进一步地,所述的多源数据包括实时数据、经常变化的结构化数据、基本不会变化的结构化数据和非结构化数据;

[0036] 所述的多源数据接入模块包括:

[0037] 实时数据接入模块,用以通过Kafka分布式消息队列将实时数据接入大数据平台;

[0038] 经常变化的结构化数据写入模块,用以通过ETL工具Kettle、增量捕获工具将经常变化的结构化数据写入大数据平台的分布式关系型数据库中;

[0039] 基本不会变化的结构化数据写入模块,用以通过Sqoop工具将基本不会变化的结构化数据写入大数据平台的分布式非关系型数据库和分布式文件系统中;

[0040] 非结构化数据写入模块,用以通过Flume工具将非结构化数据直接写入大数据平台的分布式文件系统。

[0041] 进一步地,所述的设备类型包括至少变电站、主变、线路、配变和开关,所述的电压等级范围为500kV~10kV,所述的业务领域包括台账信息、运行信息、空间信息和规划信息。

[0042] 进一步地,所述统一模型包括逻辑模型、拓扑模型和特征模型,

[0043] 所述逻辑模型为以高电压等级变电站为根节点,按设备类型分类为分支,10kV配变为叶子节点的树形数据结构,做为各业务系统电网设备的逻辑模型;

[0044] 所述拓扑模型为以电气接线关系、地理接线关系拓扑图,以及电压等级由高到底、设备类型统一的拓扑模型有向图数据结构,做为各业务系统电网设备的拓扑模型;

[0045] 所述特征模型为根据设备共有属性信息提取设备特征模型线性结构,对设备类型进行统一编码,每个设备类型建立独立的对象模型,以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类,以每类设备特有属性为子类的特征模型。

[0046] 进一步地,所述模型匹配度计算模块包括:

[0047] 逻辑模型计算模块,用以通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦,以及在确保根节点已对应的情况下,对各系统中的逻辑模型进行横

向匹配,计算各节点包含下级节点数量,并做为相应节点的属性;

[0048] 拓扑模型计算模块,用以通过有向图的拓扑排序算法计算拓扑模型中节点的前驱及后续节点关系,首先在有向图中选取一个没有前驱的顶点,将其输出从有向图中删除该顶点,并且删除以该顶点为尾的所有有向图的边;重复以上的步骤,直到图中的所有顶点均输出或是图中的顶点均没有前驱为止,确立该节点的上下文在各源系统模型中是否一致;

[0049] 特征模型计算模块,用以在对设备聚焦的基础上,通过各系统中特征模型中的属性信息,对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异。

[0050] 本发明的上述技术方案的有益效果如下:

[0051] 随着大数据技术的不断成熟,基于大数据的分布式存储和并行计算技术,为多源数据快速、智能融合提供了解决问题的信息化新途径。本发明的上述技术方案提出了一种基于多元异构模型的电网多源数据融合方法,它分为多源数据接入、元数据存储、设备统一建模及数据提取、模型匹配度计算、多源数据融合五个阶段,通过各个阶段渐进式数据处理和计算,完成多元异构数据的智能融合,通过对各源系统数据的接入、智能自匹配,最终生成电网多源数据融合库,包含统一的设备类型、设备对应关系、数据融合过程记录等内容,在后续数据的应用中,通过该库可以灵活的实现电网各业务数据的关联查询、统计,实现了设备数据库、运行数据库、图形数据库、规划数据库的“四库合一”。

[0052] 一种基于多元异构模型的电网多源数据融合方法,通过深入分析多源系统数据,建立包括各源系统数据逻辑关系、拓扑结构、特征属性等模型,通过模型遍历比对,计算模型间匹配度,根据匹配度情况进行自动关联,实现了90%以上的数据自动集成融合,并准确分析出未能融合数据明细及具体原因,按数据所属电压等级和设备所在区域以任务的方式在电网规划综合信息平台中分发给相应用户进行手工修补。既实现了各业务数据的集成融合,又对整体的融合质量进行了有效的管控。

[0053] 通过采用ETL、webservice、Sqoop增量捕获、大数据分布式存储和并行计算等多种技术手段,实现结构化、非结构化、海量历史/准实时、电网空间数据等信息的高效接入以及电网设备、运行水平、图形管理、规划方案等数据的深度融合,有效集成了PMS、GIS、EMS、智能配网监控、用电信息采集等多个电力业务系统,形成了集数据融合、存储、计算、建模、挖掘、分析及智能感知于一体的电网规划综合数据库,实现了设备数据库、运行数据库、图形数据库、规划数据库的“四库合一”,支撑了电网规划诊断分析、负荷分析预测、电网仿真计算、能源电力供需、经济技术评估等业务的开展,实现了省域电网发展专业数据资产的集中管理与高度共享,保证了全省电网数据的权威性、统一性和全面性,既可为公司发展专业全链条业务提供支撑,也可为能源主管部门提供综合信息咨询服务。

附图说明

[0054] 图1为本发明实施例提供的一种基于多元异构模型的电网多源数据融合方法的方法流程图;

[0055] 图2为本发明实施例提供的一种基于多元异构模型的电网多源数据融合系统的结构示意图;

- [0056] 图3为本发明实施例中所述拓扑模型的示意图；
[0057] 图4为本发明实施例中所述特征模型的示意图；
[0058] 图5为本发明实施例中所述模型匹配度计算的示意图；
[0059] 图6为本发明实施例在进行多源数据融合的过程图。

具体实施方式

[0060] 为能清楚说明本方案的技术特点,下面通过具体实施方式并结合其附图对本发明进行详细阐述。下文的公开提供了许多不同的实施例或例子用来实现本发明的不同结构。为了简化本发明的公开,下文中对特定例子的部件和设置进行描述。此外,本发明可以在不同例子中重复参考数字和/或字母。这种重复是为了简化和清楚的目的,其本身不指示所讨论各种实施例和/或设置之间的关系。应当注意,在附图中所图示的部件不一定按比例绘制。本发明省略了对公知组件和处理技术及工艺的描述以避免不必要地限制本发明。

[0061] 为了更好地理解本发明,下面结合附图来详细解释本发明的实施方式。实施例一

[0062] 参见图1,图1为本发明实施例提供的一种基于多元异构模型的电网多源数据融合方法的方法流程图。由图1可知,本发明实施例中的一种基于多元异构模型的电网多源数据融合方法,它包括以下过程:

[0063] 多源数据接入:采用多种技术手段进行结构化数据、非结构化数据、海量历史/准实时数据和电网空间数据的接入;

[0064] 元数据存储:对各类数据按照统一数据规范进行标准化格式存储;

[0065] 设备统一建模及数据提取:确定并统一需要进行数据融合的设备类型、电压等级范围和业务领域信息,建立多源数据的统一模型(即多元异构模型),并进行数据提取和比对计算;

[0066] 模型匹配度计算:根据多源数据的统一模型进行设备匹配度计算,通过统一模型中的属性信息对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异;

[0067] 多源数据融合:对于A类数据直接建立关联关系;对于B类数据则开发数据融合对应工具,用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系;对应C类数据则以问题的形式反馈至源系统进行数据修补,修补完成后再根据修补增量实现自动关联。

[0068] 在本发明的实施例中,所述的多源数据包括实时数据、经常变化的结构化数据、基本不会变化的结构化数据和非结构化数据;所述的多源数据接入的具体过程为:

[0069] 通过Kafka分布式消息队列将实时数据接入大数据平台;

[0070] 通过ETL工具Kettle、增量捕获工具将经常变化的结构化数据写入大数据平台的分布式关系型数据库中;

[0071] 通过Sqoop工具将基本不会变化的结构化数据写入大数据平台的分布式非关系型数据库和分布式文件系统中;

[0072] 通过Flume工具将非结构化数据直接写入大数据平台的分布式文件系统。

[0073] 在本发明的实施例中,在设备统一建模及数据提取过程中,所述的设备类型包括

至少变电站、主变、线路、配变和开关,所述的电压等级范围为500kV~10kV,所述的业务领域包括台账信息、运行信息、空间信息和规划信息。

[0074] 在本发明的实施例中,所述统一模型包括逻辑模型、拓扑模型和特征模型,

[0075] 所述逻辑模型为以高电压等级变电站为根节点,按设备类型分类为分支,10kV配变为叶子节点的树形数据结构,做为各业务系统电网设备的逻辑模型;

[0076] 所述拓扑模型为以电气接线关系、地理接线关系拓扑图,以及电压等级由高到底、设备类型统一的拓扑模型有向图数据结构,做为各业务系统电网设备的拓扑模型;

[0077] 所述特征模型为根据设备共有属性信息提取设备特征模型线性结构,对设备类型进行统一编码,每个设备类型建立独立的对象模型,以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类,以每类设备特有属性为子类的特征模型。

[0078] 在本发明的实施例中,所述模型匹配度计算的具体过程为:

[0079] 通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦;

[0080] 在确保根节点已对应的情况下,对各系统中的逻辑模型进行横向匹配,计算各节点包含下级节点数量,并做为相应节点的属性;

[0081] 通过有向图的拓扑排序算法计算拓扑模型中节点的前驱及后续节点关系,首先在有向图中选取一个没有前驱的顶点,将其输出从有向图中删除该顶点,并且删除以该顶点为尾的所有有向图的边;重复以上的步骤,直到图中的所有顶点均输出或是图中的顶点均没有前驱为止,确立该节点的上下文在各源系统模型中是否一致;

[0082] 在对设备聚焦的基础上,通过各系统中特征模型中的属性信息,对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异。

[0083] 将电网多源数据融合过程分为多源数据接入、元数据存储、设备统一建模及数据提取、模型匹配度计算、多源数据融合五个阶段,通过各个阶段渐进式数据处理和计算,完成多元异构数据的智能融合,通过对各源系统数据的接入、智能自匹配,最终生成电网多源数据融合库,包含统一的设备类型、设备对应关系、数据融合过程记录等内容,在后续数据的应用中,通过该库可以灵活的实现电网各业务数据的关联查询、统计,实现了设备数据库、运行数据库、图形数据库、规划数据库的“四库合一”。

[0084] 通过采用ETL、webservice、Sqoop增量捕获、大数据分布式存储和并行计算等多种技术手段,实现结构化、非结构化、海量历史/准实时、电网空间数据等信息的高效接入以及电网设备、运行水平、图形管理、规划方案等数据的深度融合,有效集成了PMS、GIS、EMS、智能配网监控、用电信息采集等多个电力业务系统,形成了集数据融合、存储、计算、建模、挖掘、分析及智能感知于一体的电网规划综合数据库,实现了设备数据库、运行数据库、图形数据库、规划数据库的“四库合一”,支撑了电网规划诊断分析、负荷分析预测、电网仿真计算、能源电力供需、经济技术评估等业务地开展,实现了省域电网发展专业数据资产的集中管理与高度共享,保证了全省电网数据的权威性、统一性和全面性,既可为公司发展专业全链条业务提供支撑,也可为能源主管部门提供综合信息咨询服务。实施例二

[0085] 参见图2,图2为本发明实施例提供的一种基于多元异构模型的电网多源数据融合系统的结构示意图。由图2可知,本发明实施例中的一种基于多元异构模型的电网多源数据融合系统,它包括:

[0086] 多源数据接入模块,用以采用多种技术手段进行结构化数据、非结构化数据、海量历史/准实时数据和电网空间数据的接入;

[0087] 元数据存储模块,用以对各类数据按照统一数据规范进行标准化格式存储;

[0088] 设备统一建模及数据提取模块,用以确定并统一需要进行数据融合的设备类型、电压等级范围和业务领域信息,建立多源数据的统一模型,并进行数据提取和比对计算;

[0089] 模型匹配度计算模块,用以根据多源数据的统一模型进行设备匹配度计算,通过统一模型中的属性信息对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异;

[0090] 多源数据融合模块,用以对于A类数据直接建立关联关系;对于B类数据则开发数据融合对应工具,用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系;对应C类数据则以问题的形式反馈至源系统进行数据修补,修补完成后再根据修补增量实现自动关联。

[0091] 在本发明的实施例中,所述的多源数据包括实时数据、经常变化的结构化数据、基本不会变化的结构化数据和非结构化数据;

[0092] 所述的多源数据接入模块包括:

[0093] 实时数据接入模块,用以通过Kafka分布式消息队列将实时数据接入大数据平台;

[0094] 经常变化的结构化数据写入模块,用以通过ETL工具Kettle、增量捕获工具将经常变化的结构化数据写入大数据平台的分布式关系型数据库中;

[0095] 基本不会变化的结构化数据写入模块,用以通过Sqoop工具将基本不会变化的结构化数据写入大数据平台的分布式非关系型数据库和分布式文件系统中;

[0096] 非结构化数据写入模块,用以通过Flume工具将非结构化数据直接写入大数据平台的分布式文件系统。

[0097] 在本发明的实施例中,所述的设备类型包括至少变电站、主变、线路、配变和开关,所述的电压等级范围为500kV~10kV,所述的业务领域包括台账信息、运行信息、空间信息和规划信息。

[0098] 在本发明的实施例中,所述统一模型包括逻辑模型、拓扑模型和特征模型,

[0099] 所述逻辑模型为以高电压等级变电站为根节点,按设备类型分类为分支,10kV配变为叶子节点的树形数据结构,做为各业务系统电网设备的逻辑模型;

[0100] 所述拓扑模型为以电气接线关系、地理接线关系拓扑图,以及电压等级由高到底、设备类型统一的拓扑模型有向图数据结构,做为各业务系统电网设备的拓扑模型;

[0101] 所述特征模型为根据设备共有属性信息提取设备特征模型线性结构,对设备类型进行统一编码,每个设备类型建立独立的对象模型,以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类,以每类设备特有属性为子类的特征模型。

[0102] 在本发明的实施例中,所述模型匹配度计算模块包括:

[0103] 逻辑模型计算模块,用以通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦,以及在确保根节点已对应的情况下,对各系统中的逻辑模型进行横向匹配,计算各节点包含下级节点数量,并做为相应节点的属性;

[0104] 拓扑模型计算模块,用以通过有向图的拓扑排序算法计算拓扑模型中节点的前驱

及后续节点关系,首先在有向图中选取一个没有前驱的顶点,将其输出从有向图中删除该顶点,并且删除以该顶点为尾的所有有向图的边;重复以上的步骤,直到图中的所有顶点均输出或是图中的顶点均没有前驱为止,确立该节点的上下文在各源系统模型中是否一致;

[0105] 特征模型计算模块,用以在对设备聚焦的基础上,通过各系统中特征模型中的属性信息,对每类设备进行细化匹配,各类属性一致的设备相互匹配度大于80%则记为A类、相互匹配度为60%~80%记为B类、相互匹配度小于60%则记为C类,并记录设备特征属性的差异。

[0106] 本发明实施例中的电网多源数据融合方法主要分为多源数据接入、元数据存储、设备统一建模及数据提取、模型匹配度计算和多源数据融合共五个阶段,通过各个阶段渐进式数据处理和计算,完成多元异构数据的智能融合,实现了设备数据库、运行数据库、图形数据库、规划数据库的“四库合一”。以下对各个阶段进行详细说明。

[0107] 1、基于多种技术手段的多源数据接入

[0108] 采用Kafka、Sqoop、增量捕获工具等多种技术手段,实现结构化、非结构化、海量历史/准实时、电网空间数据(全量和增量捕获)接入,对各类数据按照统一数据规范进行标准化格式存储,依据应用需求存储在分布式关系型数据库、分布式非关系型数据库和分布式文件系统中。

[0109] 1) 实时数据(如:负荷数据、用户用电数据、设备状态监测数据等)通过Kafka分布式消息队列接入大数据平台。

[0110] 2) 结构化数据(经常变化的事务型操作数据)通过ETL工具Kettle、增量捕获工具,将数据写入大数据平台的分布式关系型数据库中。

[0111] 3) 结构化数据(基本不会变化的数据。如历史数据)通过Sqoop工具,将数据写入大数据平台的分布式非关系型数据库和分布式文件系统中。

[0112] 4) 非结构化数据(如日志文件等)通过Flume工具将数据直接写入大数据平台的分布式文件系统。

[0113] 2、基于大数据分布式技术的元数据存储

[0114] 1) 关系型数据库满足计算结果数据、地理空间基础数据、主数据等具有二维技术特性的数据存储需求。

[0115] 2) 分布式文件系统存储非结构化相关数据。HDFS文件系统是建立在低成本X86硬件上的分布式文件系统集群,采用主从结构,由主节点负责分布式文件系统的元数据管理和提供统一的命名空间,由数量众多的数据节点负责数据IO处理和计算。HDFS解决方案中,数据文件将被划分成一个或多个数据块,并分散存储在不同的数据节点上,数据块有多个冗余,以解决硬件故障导致的数据丢失问题。

[0116] 3) 分布式数据库满足具有时间序列特性、地理空间瓦片数据存储,如测点数据、负荷运行信息等。

[0117] 3、基于大数据并行计算技术实现多元异构电网设备统一模型及数据提取

[0118] 确定并统一需要进行数据融合的设备类型(变电站、主变、线路、配变、开关等)、电压等级范围(500kV~10kV)、业务领域(台账信息、运行信息、空间信息、规划信息)等信息,在此基础上分别对各业务系统数据建立以下统一模型,并通过SPARK技术进行数据提取和比对计算。

[0119] 1) 逻辑模型

[0120] 分析各源系统数据结构,生成以高电压等级变电站为根节点,按设备类型分类为分支,10kV配变为叶子节点的树形数据结构,做为各业务系统电网设备的逻辑模型,通过Spark大数据技术开发各源系统模型数据按根节点查询方法,实现各源系统逻辑模型数据的快速构建。

[0121] 2) 拓扑模型

[0122] 如图3所示,分析各源系统数据,生成电气接线关系、地理接线关系拓扑图,电压等级由高到底、设备类型统一的拓扑模型有向图数据结构,做为各业务系统电网设备的拓扑模型,通过Spark大数据技术开发各源系统拓扑模型数据按起始节点查询方法,实现各源系统拓扑模型数据的快速构建。

[0123] 3) 特征模型

[0124] 如图4所示,分析各源系统数据,根据设备共有属性信息提取设备特征模型线性结构,对设备类型进行统一编码,每个设备类型建立独立的对象模型,以数据源系统代码、ID、设备名称、设备类型、电压等级共有属性为基类,以每类设备特有属性为子类的特征模型,通过Spark大数据技术开发各源系统特征模型数据按父类属性查询方法,实现各源系统特征模型数据的快速构建。

[0125] 4、基于多源数据统一模型的设备匹配度计算

[0126] 如图5所示,设备匹配度计算过程如下:1)通过对逻辑模型树形数据结构的广度优先遍历算法实现树节点逐个设备聚焦,由于数据量大,采用非递归模式遍历以提升算法的效率和稳定性。

[0127] 2)在确保根节点已对应的情况下,对各系统中的逻辑模型进行横向匹配,计算各节点包含下级节点数量,做为本节点属性。

[0128] 3)通过有向图的拓扑排序算法,计算拓扑模型中节点(电气设备)的前驱及后续节点关系,首先在有向图中选取一个没有前驱的顶点,将其输出,从有向图中删除该顶点,并且删除以该顶点为尾的所有有向图的边。重复以上的步骤,直到图中的所有顶点均输出或是图中的顶点均没有前驱为止,确立设备的上下文在各源系统模型中是否一致,已有模型单个设备匹配度计算中,所有系统拓扑模型能够上下文相互匹配大于80%则记为A、60%~80%则记为B、60%以下记为C,并记录在哪些系统拓扑模型统一,哪些系统拓扑模型不统一。

[0129] 4)在对设备聚焦的基础上,通过各系统中特征模型中的属性信息,对每类设备进行细化匹配,各类属性一致的设备相互匹配大于80%则记为A、60%~80%记为B、60%以下记为C,并记录设备在各系统中特征属性的差异。

[0130] 5、基于关系数据库的多源数据融合

[0131] 通过以上多元数据匹配度计算,A类数据直接建立关联关系;B类数据,开发数据融合对应工具,用户再次根据设备的融合记录、推荐对应设备及实际情况手动建立关联关系;C类数据大多属于冗余数据或需源系统修补数据,以问题的形式反馈至源系统进行数据修补,修补完成后再根据修补增量实现自动关联。

[0132] 图6为本发明实施例在进行多源数据融合的过程图,基于多元异构模型智能自匹配算法的电网多源数据融合方法通过对各源系统数据的接入、智能自匹配,最终生成电网

多源数据融合库,包含统一的设备类型、设备对应关系、数据融合过程记录等内容,在后续数据的应用中,通过该库可以灵活的实现电网各业务数据的关联查询、统计。

[0133] 本发明通过深入分析多源系统数据,建立包括各源系统数据逻辑关系、拓扑结构、特征属性等模型,通过模型遍历比对,计算模型间匹配度,根据匹配度情况进行自动关联,实现了90%以上的数据自动集成融合,并准确分析出未能融合数据明细及具体原因,按数据所属电压等级和设备所在区域以任务的方式在电网规划综合信息平台中分发给相应用户进行手工修补。既实现了各业务数据的集成融合,又对整体的融合质量进行了有效的管控。

[0134] 此外,本发明的应用范围不局限于说明书中描述的特定实施例的工艺、机构、制造、物质组成、手段、方法及步骤。从本发明的公开内容,作为本领域的普通技术人员将容易地理解,对于目前已存在或者以后即将开发出的工艺、机构、制造、物质组成、手段、方法或步骤,其中它们执行与本发明描述的对应实施例大体相同的功能或者获得大体相同的结果,依照本发明可以对它们进行应用。因此,本发明所附权利要求旨在将这些工艺、机构、制造、物质组成、手段、方法或步骤包含在其保护范围内。



图1



图2

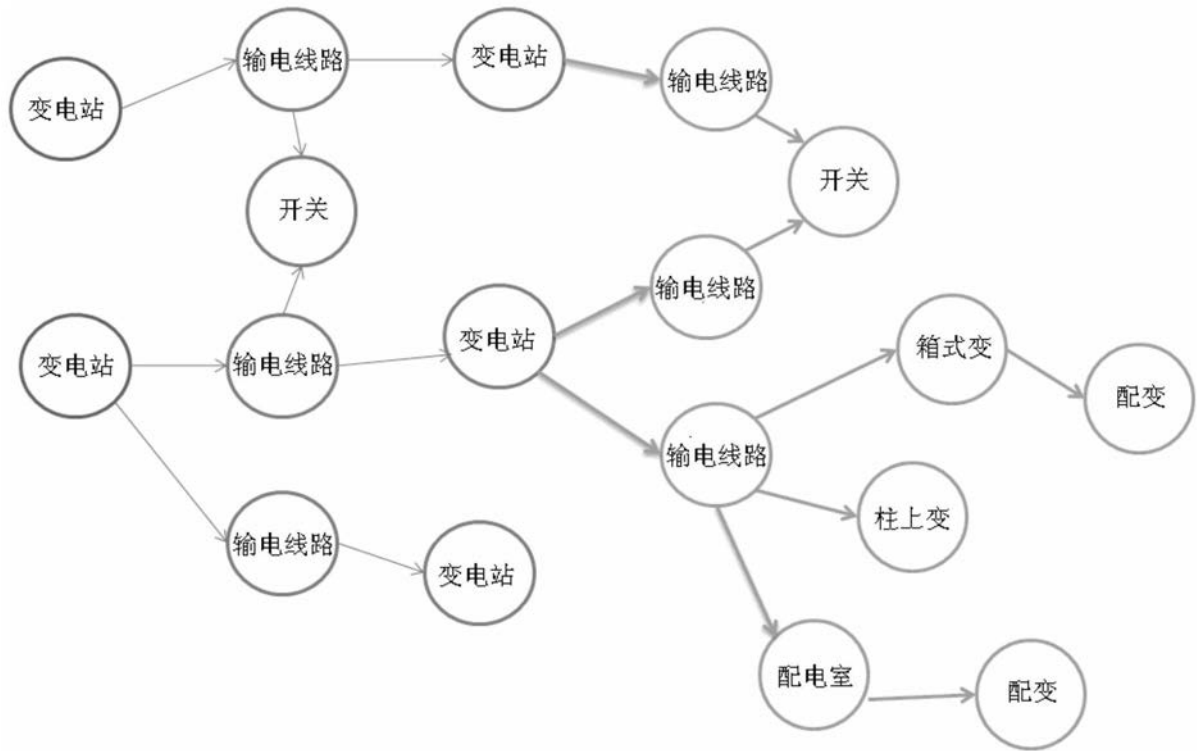


图3

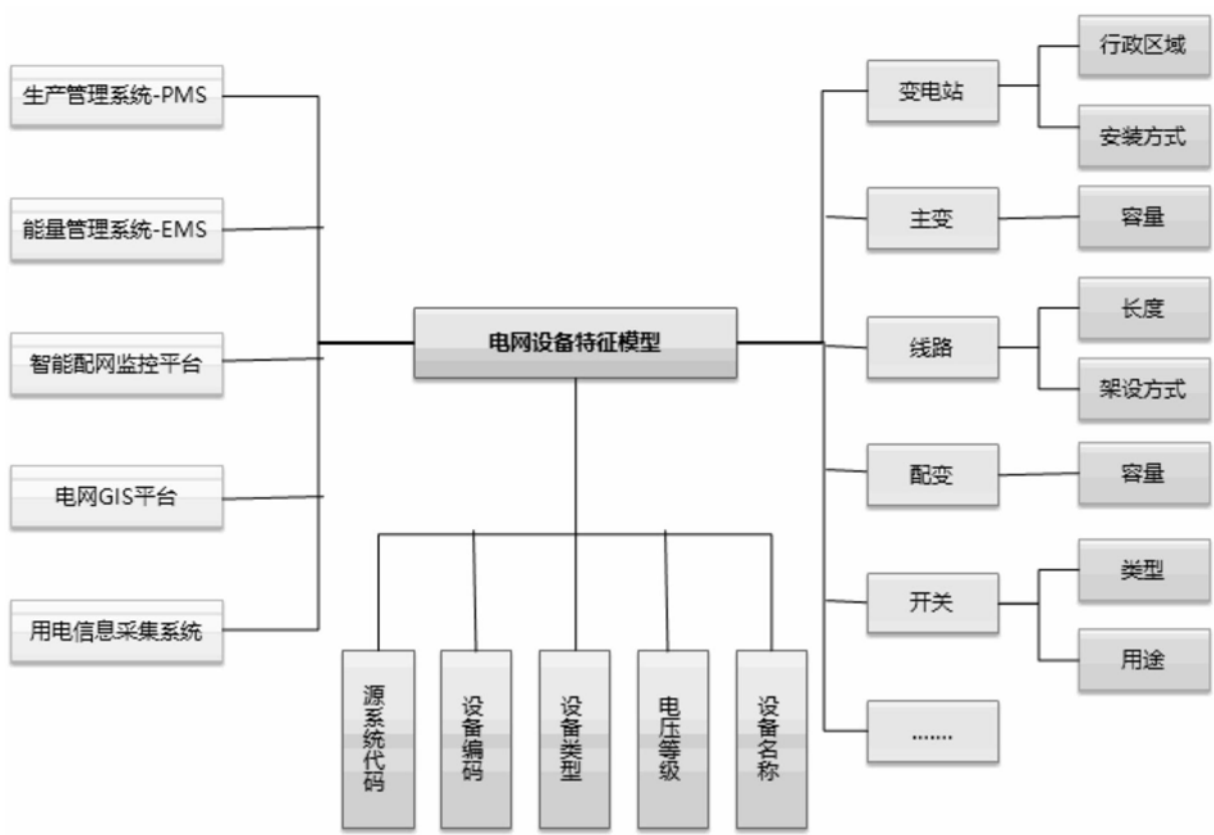


图4

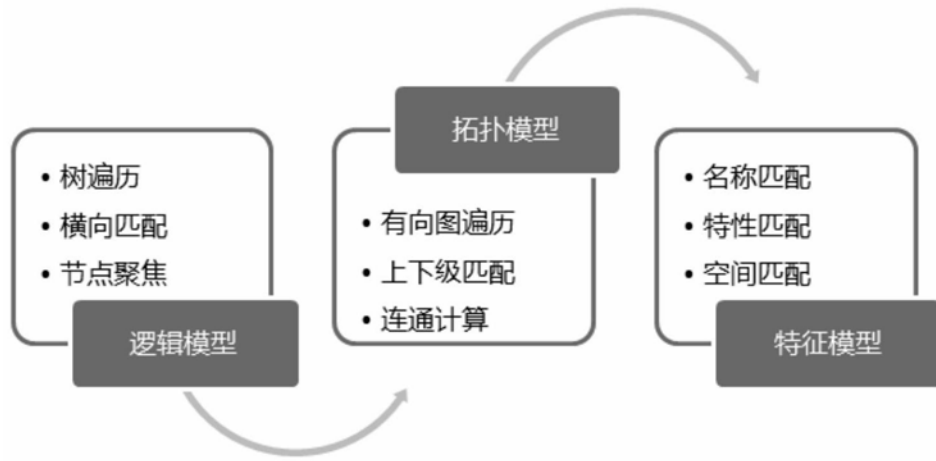


图5

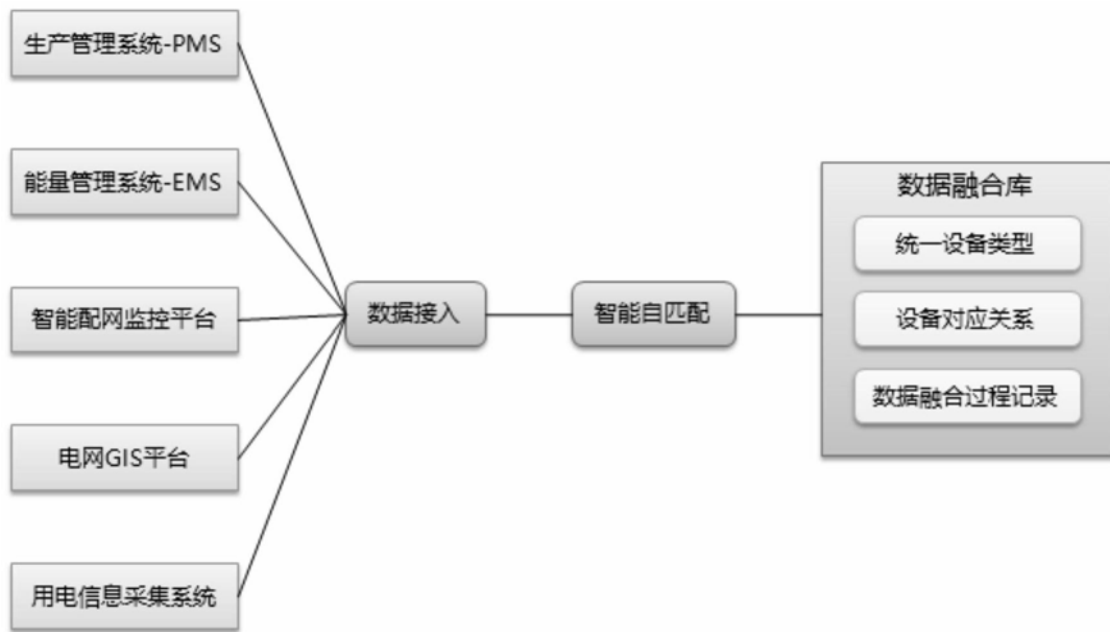


图6