

基于 Apriori 算法的电厂辅助服务考核数据分析与应用

孟科技

(徐州华润电力有限公司, 江苏 徐州 221142)

摘要:为了保障电力系统的供电安全性、稳定性和可靠性及维持电能的优质经济运行,需要发电企业、电网企业等具有调频、自动发电控制、调峰等服务。鉴于火电机组的性能差异性,可能在某一时刻不能满足部分辅助服务考核项目而被考核。调度根据规程对其基本辅助服务和有偿辅助服务相关内容进行考核和奖励。针对机组因考核被罚款的问题,本文采用 Apriori 算法对电厂侧的数据进行挖掘分析。Apriori 算法作为大数据算法其中一种方法,它是一种基于递推算法思想算法。在候选集生成和情节的向下封闭检测的两阶段时,通过频繁项集的关联规则的挖掘并获取各考核项目的内在关系。通过对电厂辅助服务考核数据的关联性进行了分析和挖掘,数据挖掘得到信息为我们做相应的优化调整和减少考核等决策提供了参考依据。

关键词:辅助服务;大数据算法;频繁项集;关联性;数据挖掘

文章编号:2096-4633(2019)09-0020-08 **中图分类号:**TM76 **文献标志码:**B

近几年来,随着社会的发展和技术的进步,电力企业为了适应社会的经济发展,正经历着一场的变革,由原先实行发电、输电、配电一体化的体制向竞争的电力市场体制转变。在这种体制下,用户的需求越来越高和同行业之间的竞争越来越激烈,电力企业为了生存下去就要生产低成本的电能,而电网需要合理输送和分配高质量电能给用户。电网依靠相关的辅助措施把优质、廉价的电能分配给用户。

在电力市场条件下,由于发电竞争的引入和电网开放,电网的运行、控制、调度和管理将更加复杂。如何保证电力生产、电网以及电力市场的安全、可靠运行,这给我们带来了巨大的挑战,这带来辅助服务问题越来越突出。在电力市场中,发电厂等不在无偿提供辅助服务,电网也会根据电力企业提供的辅助服务质量和贡献进行相关的费用考核和补偿。电力企业参与辅助服务就有可能牺牲自己的利益来满足电网的要求,为此参与企业也应获得相应回报。这就要求电力企业在技术上进行革新和改造,越来越被重视辅助服务相关内容。国内的“两个细则”对保障电力系统安全、优质、经济运行,维护电力企业合法权益,加强辅助服务管理和并网电厂考核工作,促进厂网协调发展,规范市场秩序,推进电力市场建设,提高电能质量和安全稳定运行水平均具有重要作用^[1]。

火电机组有偿调峰分为深度调峰和启停调峰。深度调峰指机组有功出力在其额定容量 50% 以下调峰方式,启停调峰是指机组在低峰时解列停机,高峰时前开机并网的调峰方式。调峰辅助服务是指通过调整发电机出力和状态以满足系统负荷的变化,确保系统频率稳定,是辅助服务的重要内容^[2]。调峰辅助服务包括基本调峰辅助服务和有偿调峰辅助服务两类。若基本调峰辅助服务总量无法满足电网调峰需求,则调用火电机组的有偿调峰辅助服务,并按相关规定进行补偿^[3]。

国外大量采用智能电表对用户的电负荷进行数据采集。他们通过智能表的抄表记录大量数据,采用大数据技术(聚类分析、灰色关联分析、决策树等)对粉盒数据进行预处理,分析每个用户负荷与天气、日类型等影响因素的密切关系,对负荷做短期的预测。国外调频辅助电力系统的调频容量需求一般由独立系统运营商负责预测,主要是根据高峰负荷的百分比来设置^[4]。

在智能电网环境下,动态可控负荷控制更加精准、集群作用显著且波动较小,可以作为需求侧的稳定资源来维持电网功率平衡并提供辅助服务,从而为电力系统运营带来显著效益。需求响应资源参与电力辅助服务市场,可以提高系统可靠性、灵活性,优化资源配置,提高用电效率,具有重要的社会效

益。充分利用负荷的实时可调节能力,发挥其响应速度快、响应效果可靠、响应成本低的优势^[5]。

随着中国制造 2025 提出,在其电力装备领域中,要求电力智能化、数据化。在当前大数据在数量上进入 PB 级时代,我们如何对收集的数据进行信息挖掘,是我们面临的一个问题。大数据不仅是信息量的捕获呈爆炸形式,还需要从大量数据中发现和分析价值。在数据挖掘算法中,Apriori 算法是一种最有影响的挖掘布尔关联规则频繁项集的算法。我们只有分析大量数据中项集之间的频繁模式、关联、相关性或因果结构,对数据进行挖掘,才能查找并发现存在于数据之间的关联信息以及数据之间的某种规律性^[6]。

1 Apriori 算法和辅助服务考核介绍

1.1 Apriori 算法介绍

我们首先从原始数据集合中找出所有的出现的频率高数据组的频繁集,其频繁性要大于等于我们预设的最小支持度,从而得到强关联规则。所找出的频繁集的规则必须满足最小支持度和最小可信度,即小于给定的最小可信度的规则的则被删除,通过使用了递归的方法最终生成所有频繁集。

Apriori 算法通常是以频集思想的递推算法为核心,首先对所有频集进行搜集,然后通过频集产生一定的强关联规则,这些规则通常要满足最小置信度和最小支持度,如果强关联规则产生的结果小于用户规定的最小置信度和最小支持度原则就会被删除,然后将剩下的频集再次进行扫描和搜索,在对空间进行收缩和压缩的同时顺便提升高频繁项集的置信度和支持度。

频繁项集挖掘的目的是找出给定数据集中挖掘出所有的频繁项集。现假设 $I = \{I_1, I_2, \dots, I_m\}$ 是一个由 m 种不重复项组成的集合; D 是数据库事务的集合,其中每条事务 T 是一个非空项集,使得 T_i 。每个事务都有一个唯一标识符 T_m 。任何一个项集都是项集 I 的子集,并且当这个项集由 k 种不同项目组成时,我们称为频繁 k 项集。设 X 是一个项集,它的支持数等于在数据库 D 中包含项集 X 的事务的总条数。如果 X 的支持数大于等于最小支持数阈值,那么 X 是一个频繁项集。Apriori 算法主要步骤是:

(1)扫描数据库 D ,计算每个项的支持数,得到

频繁 1 项集的集合;

(2)连接步:为了找出 $L_k + 1$,通过将 L_k 与自身连接产生 $C_k + 1$,其中 L_k 为频繁 k 项集的集合, $C_k + 1$ 为候选 $k + 1$ 项集的集合;

(3)剪枝步:由于任何非频繁的 k 项集一定不是频繁 $k + 1$ 项集的子集,所以当有一个候选 $k + 1$ 项集的 k 项子集不在 L_k 中,那么这个候选 $k + 1$ 项集不可能是频繁项集,通过这个先验性质实现对 $C_k + 1$ 进行剪枝;

(4)扫描数据集,确定 $C_k + 1$ 中每个候选候选 $k + 1$ 项集的支持数,从而获得所有的频繁 $k + 1$ 项集;

(5)通过重复步骤(2) - 步骤(4),直到不再产生频繁 $L_k + 1 (k \geq 2)$ 项集。

Apriori 是对现在影响比较大的数据挖掘算法,利用迭代形式,依次循环处理,直到生产最终的频繁 K 项集,整个过程需要多次扫描数据库,最后根据频繁 K 项集分析总结出强关联规则,供用户决策分析。我们还可以通过特征选择对于给定的数据集进行特征分析,找到它们与类关联的标签。可以生成具有预期最小支持和置信度的项目集生成规则。这种基于期望类标签选择元组的方法通过减少查找关联规则所涉及的迭代次数和时间来提高 Apriori 算法的效率。

1.2 关联规则基本定义

Apriori 算法是最常用的数据挖掘算法,数据挖掘的目的就是从庞大的、杂乱的、模糊的、不完整的、随机的数据中,希望通过特定的算法找出事物之间的内在联系。Apriori 算法是一种关联规则挖掘算法,关联规则可以用 $X \rightarrow Y$ 这种蕴含形式表达,其中 $X \cap Y = \emptyset$ 。而 X 、 Y 两个事物是否关联,有多大的关联性,则由支持度和置信度来判断。支持度是指包含 $X \cup Y$ 的事物占全体事物的百分比;置信度是指包含 $X \cap Y$ 的事物占项集 X 的百分比。支持度是判断关联规则 $X \rightarrow Y$ 是否成立;置信度是表示关联规则 $X \rightarrow Y$ 中, X 、 Y 相互关联的强度。Apriori 算法的目的是从数据集中找出满足最小支持度的所有项集,把其中最大的项集称为最大频繁项集,在最大频繁项集中挖掘满足最小置信度的关联规则。

关联规则挖掘是指从给定的事务数据库中,利用相应的关联规则算法,依据用户给出的最小支持

度和最小可信度,挖掘出事务项之间的强关联性规则。关联规则挖掘过程主要包含两个阶段:一是从事务数据库中找出所有的频繁项目集;二是从这些

频繁项目集中生产需要的关联规则^[7]。
支持度为几个关联数据项在所有数据 N 中同时出现的概率,如式(1)所示。

$$Support(T1,T2\cdots) = P(T1,T2\cdots) = \frac{\sum(T1,T2\cdots)}{N}$$

(1)

置信度是几个关联数据项在某一数据项出现后的概率,如式(2)所示。

$$Confidence(Tm \leftarrow T1,T2\cdots Tn) = \frac{P(T1,T2\cdots Tn)}{P(T1,T2\cdots Tm-1,Tm+1\cdots Tn)}$$

(2)

对于强关联规则,设 min_sup 是最小支持度阈值;min_conf 是最小置信度阈值。如果事务集合 T 中的关联规则 $A \rightarrow B$ 同时满足 $Support(A \rightarrow B) \geq min_sup$, $Confidence(A \rightarrow B) \geq min_conf$,则 $A \rightarrow B$ 称为 T 中的强关联规则。而关联规则的挖掘就是在事务集合中挖掘强关联规则。

关联规则能够反映事物与事物之间的相互依存性和关联性。如果通过数据的挖掘能够反映两个或多个事物之间存在一定的关联关系,就有助于进一步分析事物与事物间的关联机制,甚至可以通过其中一项事物的发生去预测其他事物的发生^[8]。

1.3 辅助服务考核介绍

随着电力行业市场化改革不断深入,发电企业面临着日益激烈的市场竞争,实时获取电厂辅助服务考核评价信息,不但给管理人员提供决策支持,而且给运维人员提供了技术支持。为了规范电力市场的公平有序和谐的发展,调度依据电厂并网运行考核及辅助服务补偿管理,采取的相关的经济手段和技术措施,来提高电网用电质量,以及保障电力系统安全、稳定和经济运行^[9]。

由于不同的电力市场具有不同的历史背景、电源结构、负荷分布和负荷特性,所需的辅助服务种类和数量也不同,所以没有一种标准的辅助服务市场可以适用于所有的电力市场。甚至在同一个电力市场的不同时期中,所需要的辅助服务也会随着市场的变化而变化,这种变化包括上述电源和电网结构的变化、电力供需形式的变化,还包括市场运营规则的变化、监控技术的改进和市场运作过程中吸取经验的增加等^[10]。

辅助服务项目有许多项,考核项目有:调度管理、发电曲线、调差能力、非停及临检超期、AGC 性能、一次调频、电压合格率等;补偿项目有:调峰补偿、AGC 基本补偿、AGC 调用补偿、AVC 服务补偿等等。

2 Apriori 算法应用

为了简化,我们选取某一年度一季度十个电厂的在五项考核项目是否被考核这一布尔量的数据作为 Apriori 算法的研究。其五项考核项目为基础调差、AGC 可调范围、AGC 调节速率、AGC 调节精度和一次调频性能。数据如表 1(× 标识代表被考核)。

表 1 考核项目表

Tab. 1 Check list

	考核指标 T1 (基础调差)	考核指标 T2 (AGC 调节精度)	考核指标 T3 (AGC 调节速率)	考核指标 T4 (AGC 可调范围)	考核指标 T5 (一次调频性能)
电厂 1	×	×			×
电厂 2		×		×	
电厂 3		×	×		
电厂 4	×	×		×	
电厂 5	×		×		
电厂 6		×	×		
电厂 7	×		×		
电厂 8	×	×	×		×
电厂 9	×	×	×		

单从上表格中,我们仅能看到 AGC 调节精度的考核次数较多,但是各个考核指标的内部关系我们并不知道,所以我们需要进一步对数据分析。

2.1 发掘频繁集

从辅助服务考核数据表中我们统计单相候选项集 C1 的支持度如表 2。

表 2 C1 候选项集

Tab. 2 C1 Candidate item set

C1 候选项集	支持度
{ T1 }	6/9
{ T2 }	7/9
{ T3 }	6/9
{ T4 }	2/9
{ T5 }	2/9

分析 C1 表中我设定最小支持度为 2/9,对于单相数集的频繁项集 L1 则为表 3。

表 3 频繁项集 L1

Tab. 3 Frequent itemset L1

频繁项集 L1	支持度
{ T1 }	6/9
{ T2 }	7/9
{ T3 }	6/9
{ T4 }	2/9
{ T5 }	2/9

根据频繁项集 L1 产生候选集项统计 C2 相项集的支持度如表 4。

表 4 C2 候选项集

Tab. 4 C2 Candidate item set

C2 候选项集	支持度
{ T1 ,T2 }	4/9
{ T1 ,T3 }	4/9
{ T1 ,T4 }	1/9
{ T1 ,T5 }	2/9
{ T2 ,T3 }	4/9
{ T2 ,T4 }	2/9
{ T2 ,T5 }	2/9
{ T3 ,T4 }	0/9
{ T3 ,T5 }	1/9
{ T4 ,T5 }	0/9

去掉小于最小支持度的相集,得到 2 相的频繁

项集 L2 如表 5。

表 5 频繁项集 L2

Tab. 5 Frequent itemset L2

频繁项集 L2	支持度
{ T1 ,T2 }	4/9
{ T1 ,T3 }	4/9
{ T1 ,T5 }	2/9
{ T2 ,T3 }	4/9
{ T2 ,T4 }	2/9
{ T2 ,T5 }	2/9

根据频繁项集 L2 产生候选集项统计 C3 相项集的支持度如表 6。

表 6 C3 候选项集

Tab. 6 C3 Candidate item set

C3 候选项集	支持度
{ T1 ,T2 ,T3 }	2/9
{ T1 ,T2 ,T5 }	2/9
{ T1 ,T3 ,T5 }	1/9
{ T2 ,T3 ,T4 }	0/9
{ T2 ,T3 ,T5 }	1/9
{ T2 ,T4 ,T5 }	0/9

去掉小于最小支持度的相集,得到 3 相的频繁项集 L3 如表 7。

表 7 频繁项集 L3

Tab. 7 Frequent itemset L3

频繁项集 L3	支持度
{ T1 ,T2 ,T3 }	2/9
{ T1 ,T2 ,T5 }	2/9

根据频繁项集 L3 产生候选集项统计 C4 相项集的支持度如如表 8。

表 8 C4 候选项集

Tab. 8 C4 Candidate item set

C4 候选项集	支持度
{ T1 ,T2 ,T3 ,T5 }	1/9

去掉小于最小支持度的相集,得到 4 相的频繁项集 L4 为空集,没有新的频繁集,算法结束。最终得到的频繁项集 { T1 }、{ T2 }、{ T3 }、{ T4 }、{ T5 }、{ T1 ,T2 }、{ T1 ,T3 }、{ T1 ,T5 }、{ T2 ,T3 }、{ T2 ,T4 }、{ T2 ,T5 }、{ T1 ,T2 ,T3 }、{ T1 ,T2 ,T5 }。

对于项集组合有很多,其组合数为,当 n 增大时

组合数增大趋势更快,我们遍历检查及运算所有的组合就会非常耗时。从图 1 组合图中我们可以发现:当某一项集是频繁项其子集也是频繁项,当某一项集是非频繁项其超集也是非频繁项。如 T125 是频繁项,其子集 T12,T15,T25 也是频繁项;T34 是非频繁项,含有 T34 的超集 T134, T234, T345, T1234, T1345, T2345, T12345 都是非频繁项。这就是 Apriori 算法的原理,通过这个原理我们可以减少遍历的次数。

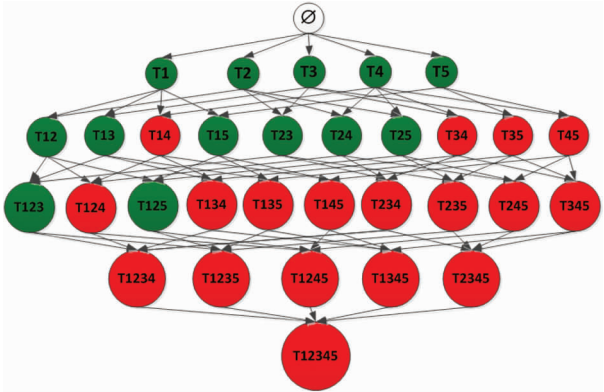


图 1 频繁集图

Fig. 1 Frequent set graph

2.2 发现关联规则

表 9 频繁集统计表

Tab. 9 Frequent set statistics table

频繁项集	支持度	规则数
{T1}	6/9	项集的长度为 1,不考虑
{T2}	7/9	项集的长度为 1,不考虑
{T3}	6/9	项集的长度为 1,不考虑
{T4}	2/9	项集的长度为 1,不考虑
{T5}	2/9	项集的长度为 1,不考虑
{T1,T2}	4/9	2
{T1,T3}	4/9	2
{T1,T5}	2/9	2
{T2,T3}	4/9	2
{T2,T4}	2/9	2
{T2,T5}	2/9	2
{T1,T2,T3}	2/9	6
{T1,T2,T5}	2/9	6

对频繁项集规则数统计得到下表,考虑项目集的长度大于 1 的频繁集。置信度是衡量关联规则准确性的指标。同时支持度反映了这种关联规则的重

要性。显然支持度越大,这种关联规则就越重要所以支持度可以剪枝处理一些毫无意义的规则。频繁项目集挖掘大都依赖于 Apriori 作为基本的修剪策略,例如通过某种安全挖掘关联的协议或增加关联规则列表等,实现中的局部修剪和联合计算。也可增加除最小支持以外的约束,删除掉枚举树中无用的或重复的节点,可以有效地削减搜索空间和减少遍历次数,实现高效挖掘强关联规则^[11]。如表 9。

通过置信度计算公式计算置信度:

规则 1:

$$T1 \rightarrow T2 \text{ 的置信度} = \frac{T12 \text{ 支持度}}{T1 \text{ 支持度}} = \frac{4/9}{6/9} = 2/3:$$

规则 2:

$$T2 \rightarrow T1 \text{ 的置信度} = \frac{T12 \text{ 支持度}}{T2 \text{ 支持度}} = \frac{4/9}{7/9} = 4/7$$

同理计算所有规则的置信度如表 10。

表 10 置信度统计表

Tab. 10 Confidence statistics

序号	规则	置信度
1	T1→T2	2/3
2	T2→T1	4/7
3	T1→T3	2/3
4	T3→T1	2/3
5	T1→T5	1/3
6	T5→T1	1
7	T2→T3	4/7
8	T3→T2	2/3
9	T2→T4	2/7
10	T4→T2	1
11	T2→T5	2/7
12	T5→T2	1
13	T12→T3	1/2
14	T3→T12	2/3
15	T13→T2	1/2
16	T2→T13	2/7
17	T23→T1	1/2
18	T1→T23	1/3
19	T12→T5	1/2
20	T15→T2	1
21	T25→T1	1
22	T1→T25	1/3
23	T5→T12	1
24	T2→T15	2/7

此时我们可以人为设置最小置信度为 $2/3$, 绿色的为大于等于置信度。如图 2。

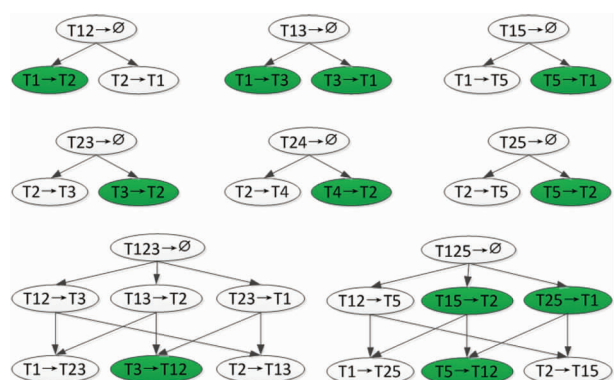


图2 强相关图

Fig. 2 Strong correlation graph

3 规则分析

在数据挖掘研究领域,对于关联分析的研究开展得比较深入,关联分析的目的是挖掘隐藏在数据间的相互关系,从原始数据中提取有意义且有用的信息的最重要任务之一,这些关系由关联规则来表示。关联性分析是综合多源、多类数据,发现和挖掘数据中潜在的相关关系,提取和挖掘数据之间的关联性充分利用大数据、发现数据要素之间的共生组合规律^[12]。关联规则是一种简单但很实用的规则,其模式属于描述型模式,挖掘关联规则的算法属于无监督学习的算法,其中 Apriori 算法是最经典的算法,该算法使用逐层搜索的迭代技术,利用已知的频繁 k 项集来生成频繁 $k+1$ 项集。算法思路直观,但实现起来需要频繁扫描数据库并进行大量计算^[13]。

因此产生的强相关联规则如下:

$T1 \rightarrow T2$, 表明基础调差被考核时,同时 AGC 调节精度有非常大的可能被考核。

$T1 \rightarrow T3, T3 \rightarrow T1$, 表明基础调差和 AGC 调节速率任一个被考核时,另一个有非常大的可能被考核。

$T5 \rightarrow T1, T5 \rightarrow T2, T5 \rightarrow T12$, 表明一次调频性能被考核时,基础调差和 AGC 调节精度其中之一或者两者一起被考核有非常大的可能性。

$T3 \rightarrow T2, T4 \rightarrow T2$, 表明 AGC 调节速率或 AGC 可调范围被考核时,同时 AGC 调节精度有非常大的可能被考核。

$T3 \rightarrow T12$, 表明 AGC 调节速率被考核时,同时基础调差和 AGC 调节精度两者一起被考核有非常大的可能性。

$T15 \rightarrow T2$, 表明基础调差和一次调频性能两者一起被考核时,同时 AGC 调节精度有非常大的可能被考核。

$T25 \rightarrow T1$, 表明 AGC 调节精度和一次调频性能两者一起被考核时,同时基础调差有非常大的可能被考核。

由此可见调频服务是辅助服务市场中的重要组成部分,调频资源根据用户要求或者 AGC 信号来自自动调整出力并以此获得报酬。

4 结论

辅助服务不仅仅是运行调度问题,而且也会随行政和经济隶属关系变化、利益主体的多元化等因素二变得更加复杂。电力需求侧管理的本质是改变用户的用电方式和行为。这两个方面分别对应的是负荷管理和能效管理。所以可以说负荷管理和能效管理都是电力需求侧管理的范畴^[14]。

Apriori 虽然是一种有效的数据挖掘算法,利用迭代形式,依次循环处理,直到生产最终的频繁 K 项集,整个过程需要多次扫描数据库,最后根据频繁 K 项集分析总结出强关联规则,供用户决策分析。但是它存在执行效率低的主要问题:多次迭代时,每次都需要遍历数据库,非常繁琐;计算候选项集时会占用很大的内存容量,寻找频繁项集需要消耗大量时间^[15]。大数据的一个基本特征是由异构和多样化维度所代表的大量数据。这是因为不同的信息收集者更喜欢他们自己的模式或协议进行数据记录,并且不同应用程序的性质也导致不同的数据表示。数据下的复杂性和关系也会随着数据量的增加而增加^[16-18]。

虽然本文的实例应用相对简单、数据量偏小,但是验证了 Apriori 算法在火电企业的应用可能性,而且为以后数据挖掘提供了一种思路和方法。通过 Apriori 算法对数据的处理,能够发现并建立潜在的数据内部关系模型。我们使用其数据内部关系模型不但可以指导我们的工作方向,而且还可以为我们的策略决策提供改善和优化的服务。

参考文献:

- [1] 王轶禹,马世俊,皮俊波,等. 关于国调直调电厂“两个细则”的讨论[J]. 电力系统自动化,2018,42(16):174-179+186.
WANG, Yiyu, MA Shijun, PI Junbo, et al. Discussion on two rules for power plants directly dispatched by state grid electric power

- dispatching and control center[J]. Automation of Electric Power Systems, 2018, 42(16): 174-179+186.
- [2] 马恒瑞,王波,高文忠,等. 考虑调频补偿效果的区域综合能源系统调频服务优化策略[J]. 电力系统自动化, 2018, 42(13): 127-134.
- MA Hengrui, WANG Bo, GAO Wenzhong, et al. Optimization strategy for frequency regulation service of regional integrated energy systems considering compensation effect of frequency regulation[J]. Automation of Electric Power Systems, 2018, 42(13): 127-134.
- [3] 李巍,袁晓婷,李俊杰,等. 基于数据挖掘的电力需求侧管理综述[J]. 电力大数据, 2018, 21(01): 10-12.
- LI Wei, YUAN Xiaoting, LI Junjie, et al. Review of electric power demand side management based on data mining[J]. Power Systems and Big Data, 2018, 21(01): 10-12.
- [4] 张全,齐红涛,刘鲲鹏,等. 国外电力企业大数据需求侧应用分析[J]. 电力大数据, 2018, 21(12): 26-31.
- ZHANG Quan, QI Hongtao, LIU Kunpeng, et al. Analysis of big data demand side application of foreign power enterprises[J]. Power Systems and Big Data, 2018, 21(12): 26-31.
- [5] 沈运帷,李扬,高赐威,等. 需求响应在电力辅助服务市场中的应用[J]. 电力系统自动化, 2017, 41(22): 151-161.
- SHEN Yunwei, LI Yang, GAO Ciwei, et al. Application of demand response in ancillary service market[J]. Automation of Electric Power Systems, 2017, 41(22): 151-161.
- [6] 郑麟. 一种直接生成频繁项集的分治 Apriori 算法[J]. 计算机应用与软件, 2014, 31(04): 297-301.
- ZHENG Lin. A divide-and-conquer Apriori algorithm directly generating frequent itemsets[J]. Computer Applications and Software, 2014, 31(04): 297-301.
- [7] 曹莹,苗志刚. 基于向量矩阵优化频繁项的改进 Apriori 算法[J]. 吉林大学学报, 2016, 54(02): 349-353.
- CAO Yin, MIAO Zhigang. Improved Apriori algorithm based on vector matrix optimization frequent items[J]. Journal of Jilin University, 2016, 54(02): 349-353.
- [8] 张延旭,胡春潮,黄曙,等. 基于 Apriori 算法的二次设备缺陷数据挖掘与分析方法[J]. 电力系统自动化, 2017, 41(19): 147-151+163.
- ZHANG Yanxu, HU Chunchao, HUANG Shu, et al. Apriori algorithm based data mining and analysis method for secondary device defects[J]. Automation of Electric Power Systems, 2017, 41(19): 147-151+163.
- [9] 刘永奇,张弘鹏,李群,等. 东北电网电力调峰辅助服务市场设计与实践[J]. 电力系统自动化, 2017, 41(10): 148-153.
- LIU Yongqi, ZHANG Hongpeng, LI Qun, et al. Design and practice of peak regulation ancillary service market for northeast China power grid[J]. Automation of Electric Power Systems, 2017, 41(10): 148-153.
- [10] 朱继忠,叶秋子,邹金,等. 英国电力辅助服务市场短期运行备用服务机制及启示[J]. 电力系统自动化, 2018, 42(17): 1-8+86.
- ZHU Jizhong, YE Qiuzi, ZOU Jin, et al. Short-term operation service mechanism of ancillary in the UK electricity market and its enlightenment[J]. Automation of Electric Power Systems, 2018, 42(17): 1-8+86.
- [11] TAMIR TASSA. Secure mining of association rules in horizontally distributed databases[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(04): 970-983.
- [12] JOSÉ MARÍA LUNA, FRANCISCO PADILLO, MYKOLA PECHENIZKIY, et al. Apriori versions based on MapReduce for mining frequent patterns on big data[J]. IEEE Transactions on Cybernetics, 2018, 48(10): 2851-2865.
- [13] 刘丽娟. 改进的 Apriori 算法的研究及应用[J]. 计算机工程与设计, 2017, 38(12): 3324-3328.
- LIU Lijuan. Research and application of improved Apriori algorithm[J]. Computer Engineering and Design, 2017, 38(12): 3324-3328.
- [14] 胡朝阳,毕晓亮,王珂,等. 促进负备用跨省调剂的华东电力调峰辅助服务市场设计[J]. 电力系统自动化, 2019, 43(05): 175-182.
- Hu Chaoyang, Bi Xiaoliang, Wang Ke, et al. Design of peak regulation auxiliary service market for east China power grid to promote inter-provincial sharing of negative reserves[J]. Automation of Electric Power Systems, 2019, 43(05): 175-182.
- [15] 陈明洁. 分布式频繁项集挖掘算法[J]. 计算机应用与软件, 2015, 32(10): 63-66.
- CHEN Mingjie. A distributed frequent itemset mining algorithm[J]. Computer Applications and Software, 2015, 32(10): 63-66.
- [16] XINDONG WU, XINGQUAN ZHU, GONG QING WU, et al. Data mining with big data[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(01): 97-107.
- [17] 赖文海,陈贤阳,明国锋,等. 基于双种群小生境差分进化算法的动态经济调度[J]. 广东电力, 2016, 29(07): 83-87+92.
- LAI Wenhai, CHEN Xianyang, MING Guofeng, et al. Dynamic economic dispatching based on dual population niche differential evolution algorithm[J]. Guangdong Electric Power, 2016, 29(07): 83-87+92.
- [18] 只健强,郭民臣,许宁,等. 机组负荷优化分配之等微增算法与动态规划算法的比较[J]. 内蒙古电力技术, 2017, 35(04): 97-100.
- ZHI Jianqiang, GUO Minchen, XU Ning, et al. Comparison between equal incremental principle and dynamic programming method in optimized unit load distribution[J]. Inner Mongolia Electric Power, 2017, 35(04): 97-100.

收稿日期: 2019-07-05

作者简介:



孟科技(1985),男,硕士,工程师,主要从事热控专业计算机系统的运行维护、DCS 协调控制等工作。

(本文责任编辑:范斌)

**Analysis and application of power plant auxiliary service assessment
data based on Apriori algorithms**

MENG Keji

(China Resources Power(XUZHOU) Co. ,Ltd. ,Xuzhou 221142 Jiangsu,China)

Abstract: In order to ensure the security, stability and reliability of power supply and maintain the high quality and economic operation of power, power generation enterprises and power grid enterprises need to have primary frequency regulation, automatic generation control, peak shaving and other services. In view of the performance differences of thermal power units, it may not be able to meet some auxiliary service assessment items at a certain time and be assessed. The dispatch shall assess and reward the relevant contents of its basic auxiliary services and paid auxiliary services according to the regulations. Aiming at the problem of unit fined for assessment, this paper uses Apriori algorithm to mine and analyze the data on the power plant side. As one of the big data algorithms, Apriori algorithm is a recursive algorithm. In the two stages of candidate set generation and downward closed detection of scenarios, the intrinsic relationship of each assessment item is obtained by mining association rules of frequent itemsets. Through the analysis and mining of the relevance of the assessment data of auxiliary services in power plants, the information obtained from data mining provides a reference basis for us to make corresponding optimization adjustments and reduce assessment decisions.

Key words: auxiliary services; big data algorithms; frequent itemsets; relevance; data mining