

doi: 10.7690/bgzd.2026.01.012

基于重构优化的电力调度数据自动化挖掘

卢 鑫¹, 康超强¹, 杜 洋², 曹泽江¹

(1. 南方电网数字电网科技(广东)有限公司项目部, 广州 510000;

2. 佰聆数据服务(广州)有限公司项目部, 广州 510000)

摘要: 为更精准地给用户 提供电力调度数据, 设计基于关联规则与重构的电力调度数据自动化挖掘方法。采用基于生成对抗网络与纵横交叉粒子群的数据重构算法, 重构电力调度数据, 补充电力调度数据中的关联缺失值; 通过基于分布式隐私保护的数据关联规则挖掘算法, 挖掘重构后电力调度数据的关联规则, 并结合多维度加密算法构建数据自动化挖掘架构, 实现电力调度数据安全挖掘。结果表明: 该方法在为 用户挖掘电力调度数据过程中, 不会出现数据丢失现象, 具有良好的自动化挖掘效果。

关键词: 关联规则; 电力调度数据; 自动化挖掘; 数据重构; 生成对抗网络; 粒子群

中图分类号: TP274 **文献标志码:** A

Automatic Mining for Power Dispatching Data Based on Reconfiguration Optimization

Lu Xin¹, Kang Chaoqiang¹, Du Yang², Cao Zejiang¹

(1. Project Department, China Southern Power Grid Digital Power Grid Technology
(Guangdong) Co., Ltd., Guangzhou 510000, China;

2. Project Department, Brilliant Data Service (Guangzhou) Co., Ltd., Guangzhou 510000, China)

Abstract: In order to provide power dispatching data to users more accurately, an automatic mining method of power dispatching data based on association rules and reconstruction is designed. A data reconstruction algorithm based on a generated antagonistic network and a crisscross particle swarm is adopted to reconstruct the power dispatching data and supplement associated missing values in the power dispatching data; Through the data association rule mining algorithm based on distributed privacy protection, the association rules of reconstructed power dispatching data are mined, and combined with multi-dimensional encryption algorithm, the data automatic mining architecture is constructed to realize the security mining of power dispatching data. The experimental results show that this method has a good automatic mining effect without data loss in the process of mining power dispatching data for users.

Keywords: association rules; power dispatching data; automatic mining; data reconstruction; generative adversarial network; particle swarm

0 引言

电网在调度过程中会产生大量的相关数据^[1-2], 而南方电网各级调度机构在对调度运行发供电与电力现货市场运营相关数据的管理、信息挖掘、发布和披露上, 存在较多的问题^[3]。其中, 最主要的问题是现有系统间仅在部分数据实现交互、统计、结算等数据挖掘, 无法满足全区域数据的动态统计等基础管理需要^[4-5]。

鉴于以上原因, 电力企业逐渐开始进行数据挖掘方法的设计, 李妍等^[6]研究数据中心网络多租户挖掘方法, 该方法通过数据感知实现多用户的数据挖掘; 但该方法在应用时, 容易出现数据丢失问题。张宇等^[7]提出了基于在线学习的挖掘数据方法, 该

方法可实现多形态数据的挖掘; 但该方法在挖掘过程中, 存在挖掘速度较慢等问题, 导致用户无法及时获取所需内容。

关联规则^[8-9]是一种反映事物与其他事物之间相关性的一种方法。通过关联规则可以展现数据之间的联系; 因此, 通过数据的关联规则挖掘, 可以挖掘出更有价值的数 据。重构对缺失数据进行补充, 通过数据重构, 可以提升数据的可靠性, 可以用到零售和市场营销、客户行为分析、金融和医疗保健及兵工领域等。笔者针对电力调度数据进行挖掘。电力调度数据指标主要包括调度运行数据和市场运营数据, 以数据列表及曲线、柱状图表等方式显示, 这些数据存储在数据库中, 支持根据关键字、时间

收稿日期: 2024-11-09; 修回日期: 2024-12-19

基金项目: 2022 年南网数研院数字电网分公司项目 (210000HK42220004)

第一作者: 卢 鑫(1984—), 男, 江西人。

等多维度进行查询。对上述数据进行挖掘,可为用户提供更全面的电力允许调度数据。

电力调度数据在复杂约束条件下,存在很强的挖掘特征模糊性,导致调度数据中的关联数据缺失。笔者结合关联规则与重构,提出电力调度数据自动化挖掘方法,满足市场运营和监管需要,为用户提供有效数据。

1 电力调度数据自动化挖掘

1.1 生成对抗网络与纵横交叉粒子群的数据重构

通过生成对抗网络与纵横交叉粒子群算法相结合的方式,构建电力调度数据重构方法,对电力调度数据中的缺失值进行重构,使后续挖掘内容更加全面。该算法的数据重构方式具体如下。

1) 通过 Wasserstein 散度生成对抗网络获取生成器与判别器的参数。图 1 为该网络的基本结构。

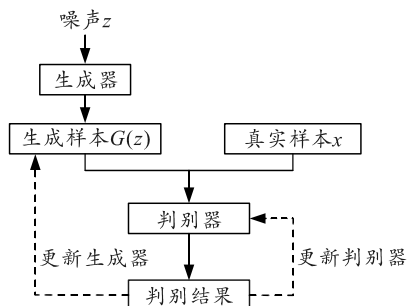


图 1 Wasserstein 散度生成对抗网络基本结构

2) 当利用 Wasserstein 散度生成对抗网络进行电力调度数据训练后,即可生成符合数据分布规律的样本为 $G(z)$ 。通过二值掩码矩阵 M 表示缺失位置:若 M 的值为 0,说明该位置存在缺失值;若为 1,则说明不存在缺失值^[10-12]。针对完整的测试样本 x ,通过 M 与 x 的哈达玛积运算($M \cdot x$),即可构成包含缺失值的样本 x' 。

3) 为保障重构后的数据样本能够更接近真实数据样本,需对电力调度数据重构过程进行合理约束,具体约束条件如下:

① 在挑选重构备选样本时,需符合原始训练样本的数据分布,因此在重构之前,需将备选样本输入到判别器中^[13-14],判定该样本是否为真实样本,设备选样本真实损失为 L_1 :

$$L_1 = D(G(z)). \quad (1)$$

式中 $D()$ 为判别器的输出。

② 从备选样本中的相应位置截取内容,以填补 x' 的缺失值,并将重构后的内容输入到判别器中,分析重构结果是否为真实样本,设重构样本为 Y ,

该样本真实损失为 L_2 ,则存在:

$$L_2 = D(Y) - L_1. \quad (2)$$

③ 通过欧氏距离判断备选样本与缺失值样本的上下文损失为 L_s :

$$L_s = d(x', L_2). \quad (3)$$

式中 $d(\bullet)$ 为变量间的欧氏距离。

通过备选样本真实损失 L_1 、重构样本真实损失为 L_2 、备选样本与缺失值样本的上下文损失 L_s 的计算,可将总损失设为 L_{total} :

$$L_{total} = L_s + \lambda_1 L_1 + \lambda_2 L_2. \quad (4)$$

式中通过 λ_1 、 λ_2 实现损失平衡。

4) 采用纵横交叉粒子群算法,对生成器的输入噪声 z 进行优化,获取最佳电力调度数据备选样本 $G'(z)$,此时,可得到重构后的电力调度数据样本:

$$X = \frac{x' + (1 - M) \cdot G'(z)}{L_{total}}. \quad (5)$$

1.2 数据挖掘过程的框架

笔者将 1.1 节的数据重构结果作为自动化挖掘内容,结合多维信息加密技术,对电力调度数据关联规则进行挖掘,设计基于维加密技术的数据挖掘架构,具体如图 2 所示。

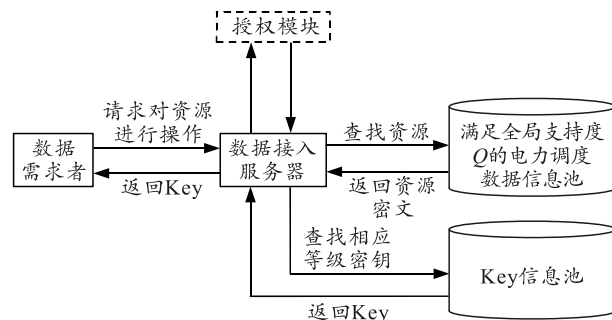


图 2 数据挖掘架构

通过 VS2010 环境搭建包含用户端与服务器端的数据自动化挖掘架构,以实现数据的安全挖掘。在图 2 中,经挖掘得到的电力调度数据关联规则存储在数据信息池中,用户与服务器之间通过 UDP 实现通信,当用户发出挖掘请求时,服务器从数据信息池里查找相应资源,同时在 Key 信息池中查找对应密钥,当密钥验证正确时,即可向用户挖掘所需资源。

通过如下步骤实现自动化挖掘架构的数据写入与读取阶段:

1) 数据读取步骤。

① 用户向服务器发送数据读取请求及其属性证书。

② 服务器查看传输来的属性证书，按照授权策略评估用户是否存在权限，若用户有权限，服务器则可根据权限赋予用户相应等级的密钥，若用户无权限，则提示用户“无此权限”并结束挖掘。

③ 用户通过私钥解密服务器提供的密钥，并按照服务器返回信息确定数据等级，利用服务器提供密钥解密密文，并对数据进行完整性验证，完成验证后，用户获取自身需求的挖掘内容。

2) 写入数据步骤。

① 用户判断写入数据类型，根据输入内容的前 3 个字符确定所属数据类型。

② 用户向服务器发送数据类型、安全登记以及自身的属性证书。

③ 服务器评估是否有权限：若有权限，则将相应密钥提供给用户；若无权限，则结束数据写入。

④ 用户通过私钥解密服务器提供的密钥，并确定数据加密等级，通过多维加密算法，加密上传内容，同时通过 MD5 算法，生成电力调度数据摘要，之后将摘要与密文同时发送至服务器端。

服务器端主要负责密钥的生成与管理，而用户端可以实现数据的加密解密以及摘要的生成。通过数据读取与写入，可以实现电力调度数据关联规则的自动化挖掘。具体挖掘步骤如图 3 所示。

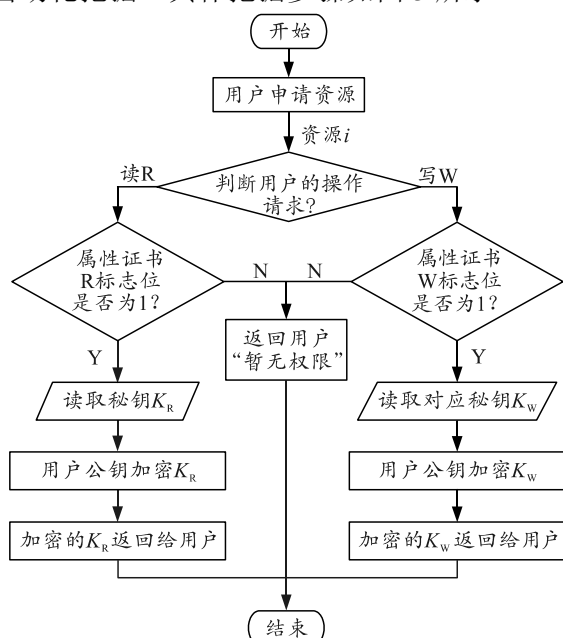


图 3 数据挖掘步骤

1.3 数据挖掘算法的设计

当完成电力调度数据重构后，通过 Apriori 挖掘算法，挖掘电力调度数据中的关联规则，使其具有更高的挖掘价值。笔者对 Apriori 算法^[15]进行优化，

向其中引入第三方 S ，使其在挖掘数据时不直接挖掘电力调度系统的数据库，从而保障数据挖掘的安全性。

通过图 4 架构表示电力调度数据的关联规则挖掘模型，该架构由数据库 DB 、挖掘者 $Miner$ 以及第三方 S 3 部分组成，其中，第三方 S 共存在 n 个站点，即表示为 $S=X\{S_1, S_2, \dots, S_n\}$ ，每个站点中包含电力调度数据，每个站点中的数据库表示为 $\{DB_1, DB_2, \dots, DB_n\}$ ；挖掘者 $Miner$ 负责数据挖掘工作；而第三方 S 负责完成计算与交互。

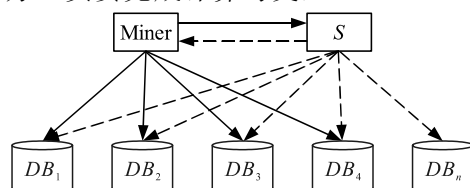


图 4 电力调度数据的关联规则挖掘基本架构

基于关联规则挖掘基本架构，研究关联规则挖掘算法挖掘水平分布方向上的电力调度数据。在电力领域中，电力调度数据具有水平分布的特点，即在一定时间范围内，电力调度数据的水平分布通常呈现出相对稳定的趋势；但电力调度数据的水平分布特点可能因具体时间段、地区和季节等因素而异，所以使用关联规则作为一种辅助手段，通过输入水平分布的电力调度数据库，在电力调度数据挖掘中帮助分析和发现不同数据项之间的关联性，求取电力相关数据的全局支持度，以深入理解数据的水平分布特点，为后期实现电力调度精准挖掘提供基础参考。

该算法通过输入水平分布的数据库，即可挖掘得到电力调度数据的关联规则。在通过该算法进行挖掘时，需要做出如下假设： S 不挖掘数据库，且 S 负责全部计算； $Miner$ 同样不挖掘数据库，但 $Miner$ 用于实现数据挖掘，并向各个站点进行汇报；而数据库则分布在 n 个站点之间，每个站点不存在互通。

针对水平分布的数据库，结合图 4 中的关联规则挖掘模型，在数据库 DB 中的某个现有节点上增加一个 S' 节点，扩展数据库 DB 中原有节点的数据类型、地理位置、时间范围等方面信息内容，较为全面的挖掘算法所能考虑的变量范围，使算法能够发现更为准确的关联规则。以此完成水平分布数据库的电力调度数据关联规则挖掘。该挖掘算法具体步骤如下：

输入：经重构后的电力调度数据水平分布数据库 $DB'=S'\{DB_1, DB_2, \dots, DB_n\}$ ；最小支持度

min_sup。

输出：电力调度数据关联规则。

1) Miner 向每个 S_i 提供一个随机数 R_i 。

2) S_i 对其中 2 个值进行计算： $S_{ik1}=F_{ik}+R_i$ ， $S_{ik2}=F_{ik}-R_i$ ，其中， F_{ik} 表示 k -项集的频繁度，且 $F_{ik}=DB'/k$ 。

3) 构建 2 个数据库，将每个项集的频繁度 F_{ik} 存放在数据库内。

4) 将每个站点的首个数据库传输给 S ，并将第 2 个数据库传输给 S' ，完成这一操作后， S 与 S' 即具备了包含噪声的项集支持数，通过这种形式，可以使得数据挖掘更具有私密性。

5) 挖掘者 Miner 采用 Apriori 算法，对 S 与 S' 进行数据挖掘处理。若 Miner 需要计算某个频繁项集的支持度时，可向 S 与 S' 发送请求，同时将随机数 R 发送给 S ； S 对 $\sum S_{ik1}$ 进行计算，并将计算结果与 R 相加后传输给 S' 。 S' 对 $\sum S_{ik2}$ 进行计算，并将计算结果与接收到 S 中的值进行相加，相加后的结果传输至 Miner，Miner 对结果除以 2，即除以数据库中的总事件数，可获取该项集的全局支持度 $Q=(\sum S_{ik1}+\sum S_{ik2})/2$ 。通过这一方式，实现电力调度数据关联规则挖掘。

2 实验分析

为了贯彻落实电力市场环境下的能源监管机构对南方电网各级调度机构的信息披露要求，将笔者设计的电力调度数据自动化挖掘方法应用至南方电网下属调度机构及其调度管辖的发电厂等运行维护单位，对其调度数据进行自动化挖掘。挖掘使用的相关数据包含南方电网发电、输电、变电、配电、用电各环节，对这些环节采集到的数据进行挖掘，验证本文中方法的应用效果。实验的具体步骤为：

1) 利用 Wasserstein 散度生成对抗网络对采集到的数据进行电力调度数据训练，生成样本 $G(z)$ ；

2) 将样本 $G(z)$ 输入到判别器中通过 L_{total} 判定该样本是否为真实样本；

3) 通过纵横交叉粒子群算法选择最佳的电力调度数据备选样本，进而得到重构原始的电力调度数据样本 X ；

4) 利用 VS2010 环境构建包含用户端与服务器端的数据自动化挖掘架构，以确保数据的安全挖掘；

5) 借助水平分布数据库和关联规则挖掘模型，输入经重构后的电力调度数据水平分布数据库 DB，

输出电力调度数据的关联规则。

当用户提出电力调度数据挖掘申请时，该用户申请挖掘广东省、广西省、云南省的电力平衡安排数据，分析应用本文中方法，能否快速实现挖掘，挖掘结果如表 1 所示。

表 1 电力平衡安排数据挖掘结果显示

数据指标	电力平衡安排/MW			数据挖掘所需时间/s
	广东省	广西省	云南省	
可见负荷	200	300	400	0.1
同比昨日负荷	200	300	400	0.2
装机容量	200	300	400	0.1
可调出力	200	300	400	0.1
送端西电电力	200	300	400	0.2
送端网外电力	200	300	400	0.1
总可调出力	200	300	400	0.1
最小发电侧备用	200	300	400	0.1
错峰情况	200	300	400	0.2
平衡盈亏	200	300	400	0.1

根据表 1 可知：当用户提出挖掘需求时，本文中方法能够快速实现相应内容的挖掘，可为用户提供精准、有效的电力调度数据；针对不同内容进行挖掘时，本文中方法均能够在 0.2 s 内提供挖掘内容；因此，本文中方法具有较高水平的自动化挖掘能力，可随时帮助用户找出所需挖掘数据。

分析不同用户在发出挖掘请求时，本文中方法所提供的挖掘内容，以此验证本文中方法能否实现自动化挖掘，具体结果如表 2 所示。

表 2 本文中方法数据挖掘情况分析

用户列表	挖掘申请内容	实际挖掘结果	是否正确
用户 1	全网最大负荷/MW	100	是
用户 2	广东省日发售电量/MWh	200	是
用户 3	全网错峰电量/MWh	100	是
用户 4	深圳能耗双控情况/MWh	800	是
用户 5	贵州送西电日实际电量/MWh	700	是
用户 6	云南省水电装机容量/MW	400	是
用户 7	广西省水电发电量/MWh	300	是
用户 8	广东省煤电出力/MW	200	是
用户 9	广西省煤电年累计发电量/MWh	300	是
用户 10	云南省气电发电量/MWh	400	是

根据表 2 可知：当不同用户提出不同内容的挖掘请求时，本文中方法均能够及时挖掘相应内容，且在挖掘过程中，未出现数据异常现象，挖掘结果全部符合实际电力调度数据；因此，本文中方法在进行数据挖掘时可有效保障数据的完整性，在较大程度上避免了数据丢失问题。

应用本文中方法，对电力调度数据的年累计发售电量与某时段的全网电负荷情况进行挖掘，分析本文中方法的挖掘结果是否完善，分析结果如图 5 所示。

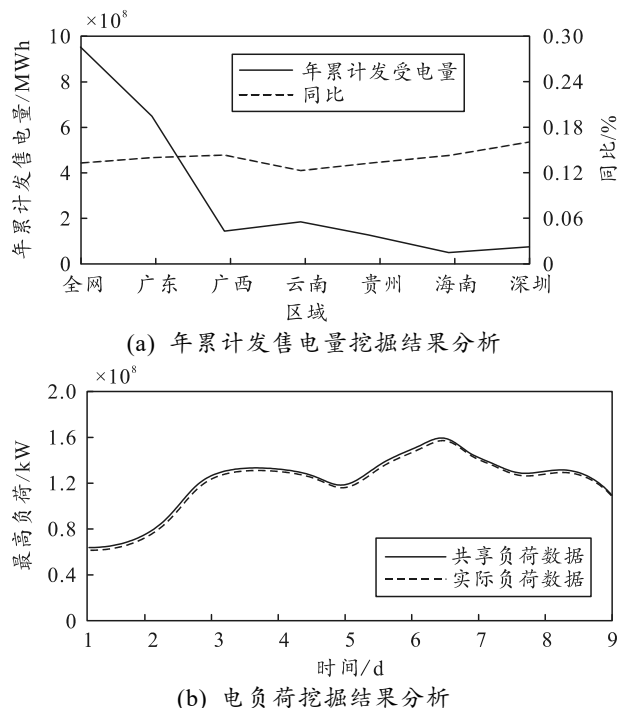


图 5 电力调度数据挖掘结果分析

根据图 5 可知：本文中方法在挖掘年累计发售电量时，能够精准提供每一城市的具体数据，同时，当进行电负荷数据挖掘时，本文中方法挖掘的数据与实际数据基本一致，未出现数据异常与数据丢失现象；因此，本文中方法具有较高性能的自动化挖掘能力。

3 结论

该研究基于关联规则与重构的电力调度数据自动化挖掘方法，结合关联规则挖掘算法与数据重构算法，对电力调度数据进行有效挖掘，为用户提供合理的电力调度数据，并通过实验验证该方法的合理性。实验结果表明：本文中方法对南方电网调度数据具有较好的挖掘效果，未出现数据丢失和异常的情况，对促进不同地区间、不同电力企业间、不同部门间对电力行业的信息挖掘和协同工作具有重要意义。在未来的研究中，考虑将所提算法应用到兵工领域中，帮助军方做出合理的电力调度决策，并及时发现和解决潜在的问题，确保军事设施和装备的正常运转，为提供有关军事装备、作战行动和军事战略等方面的洞察和决策支持。

参考文献：

- [1] 傅世元, 高欣, 张浩, 等. 基于元学习动态选择集成的电力调度数据异常检测方法[J]. 电网技术, 2022, 46(8): 3248-3261.
- [2] 屈志坚, 洪应迪, 王子潇. 配电网调度监测数据的多线程集群共享内存折叠压缩方法[J]. 中国电机工程学报, 2021, 41(3): 921-932.
- [3] 刘晶, 董志红, 张喆语, 等. 基于联邦增量学习的工业物联网数据共享方法[J]. 计算机应用, 2022, 42(4): 1235-1243.
- [4] 孟令雯, 张锐锋, 李鑫卓, 等. 基于机器学习的变电站设备异常状态数据清洗[J]. 电力系统及其自动化学报, 2021, 33(12): 79-86.
- [5] 李铁成, 任江波, 刘清泉, 等. 基于深度学习的智能录波器配置数据自动化映射方法[J]. 电测与仪表, 2022, 59(9): 76-83.
- [6] 李妍, 郭得科, 曹晓丰, 等. 应用感知的数据中心网络多租户共享方法[J]. 计算机学报, 2021, 44(7): 1363-1377.
- [7] 张宇, 刘威, 邵良杉. 面向分布式在线学习的共享数据方法[J]. 控制与决策, 2021, 36(8): 1871-1880.
- [8] 王宏刚, 王一蓉, 赵晓龙. 基于用户综合用能信息的用能行为异常识别系统设计[J]. 自动化仪表, 2022, 43(11): 99-104.
- [9] 甘昕艳, 唐晓年. 基于 CNN 的时序数据关联规则挖掘模型[J]. 计算机仿真, 2021, 38(3): 282-285, 326.
- [10] 林超, 何德彪, 黄欣沂. 基于区块链的电子医疗记录安全共享[J]. 计算机应用, 2022, 42(11): 3465-3472.
- [11] 蒋家昊, 张璇, 邓宏镜, 等. 基于区块链的多部门数据共享访问控制流程建模[J]. 计算机集成制造系统, 2022, 28(10): 3202-3211.
- [12] 陈丽燕, 吕光金, 芮廷先. 基于区块链的商业数据共享模型[J]. 计算机工程与设计, 2021, 42(5): 1257-1263.
- [13] 刘雪娇, 殷一丹, 陈蔚, 等. 基于区块链的车联网数据安全共享方案[J]. 浙江大学学报(工学版), 2021, 55(5): 957-965.
- [14] 杨业平, 林德威, 黄芳芳, 等. 基于区块链的物联网安全数据共享系统[J]. 福州大学学报(自然科学版), 2021, 49(6): 739-746.
- [15] 吴昊, 刘钊, 顾进广. 基于哈希表与十字链表存储的 Apriori 算法优化[J]. 计算机应用与软件, 2022, 39(7): 247-255.