

基于差分隐私与联邦学习的多源异构电力营销数据一致性校验算法

戴媛媛¹ 司健宏¹ 梁锐凯¹
DAI Yuanyuan SI Jianhong LIANG Ruikai

摘要

为实现对电力营销数据中异常数据的检测,并降低校验过程中的数据损失,文章基于差分隐私与联邦学习技术,以多源异构电力营销数据为例,设计其一致性校验算法。基于联邦学习算法,进行多源异构电力营销数据的融合处理;将差分隐私理论引入数据一致性判定矩阵,在保证数据保密性的同时,实现高效验证;利用所构建的校验矩阵,结合数据间的属性对应关系,进行数据一致性校验。实验结果表明,应用文章所设计的方法,可以实现对所有异常数据的检测,并召回所有检测到的异常数据,能使多源异构电力营销数据一致性校验的损失度快速收敛,并在收敛后保持在较低水平。

关键词

差分隐私;一致性校验算法;电力营销数据;多源异构;联邦学习

doi: 10.3969/j.issn.1672-9528.2026.05.034

0 引言

电力营销数据呈现出爆炸式增长与多源异构化的发展趋势。这些数据广泛分布于发电企业、电网公司、售电主体及用户侧智能终端等多个参与方,涵盖了用电量、电价、负荷曲线、合同信息等多模态内容。高质量的数据是实现精准负荷预测、电价政策制定、反窃电分析及用户服务质量提升的核心基础,而数据一致性则是保障其质量与可信度的关键前提。然而,由于数据来源多样、格式规范不一、传输环节复杂,以及各主体因商业机密与用户隐私保护需求而形成的“数据孤岛”效应,使得传统依赖于数据集中聚合的一致性校验方法面临巨大挑战。

孔祥磊等^[1]利用图论、李雅普诺夫稳定性等方法,设计了具有方向性约束的分布式控制器,并针对多机器人编队、传感器网络覆盖等应用场景进行了数值模拟。但定向约束矩阵构造需要事先确定主体之间的定向关系,很难自适应于动态环境中的实时调节。卫泽晨等^[2]通过构造多层次默克尔树、哈希校验、分级存储等手段,并结合区块链的抗篡改性能,实现版本数据的可靠溯源和整体一致性验证。然而,区块链一致性机制的引入增加了系统版本更新的时延,难以满足电网对实时性的严格要求。

为解决现有方法在应用中的缺陷与不足,提高数据的质量,本文将基于差分隐私与联邦学习的应用,以多源异构电力营销数据为例,对其一致性校验算法展开设计。

1 基于联邦学习的多源异构电力营销数据融合

在电力市场中,由于数据分布于不同地域、不同业务系统,呈现出多源异质特性,使得数据的一致性检验与综合利用面临着较大挑战。联合学习是解决该问题的有效方法之一,该方法可以在保证用户隐私的同时,将多源异构的电力市场数据进行融合。

在此过程中,利用中心服务器进行初始化计算,设置全局模型、局部模型、条件生成对抗网络(CGAN)模型及相应的学习率等参数,并在此基础上随机选取参与当前一轮训练的发电单元,通过广播方式将CGAN参数传播给每一个发电单元。根据当地现有的用电数据,包括客户用电信息、用电缴费记录和业务处理记录等,对CGAN网络进行训练^[3]。在训练过程中,从局部数据集中随机生成噪声向量,通过网络生成假样本,并使用判别网络对其进行分类,将生成器和判别器的参数进行反向传播,以实现CGAN网络中数据的纳什均衡^[4]。此过程用公式表示为:

$$\min V(D) = E_x [\log D(x|y)] + E_y [1 - \log D(G(x|y))] \quad (1)$$

式中: V 表示多源异构电力营销数据纳什均衡处理, D 表示噪声向量, E_x 、 E_y 表示虚拟样本, G 表示分类函数, x 、 y 表示多源异构电力营销数据集。联邦学习算法通过对生成器与判别器的对抗训练,使生成器能够产生更加真实的样本,从而更好地反映出数据的整体分布特性。

中心服务器根据每个单元学习到的数据集中多个局部用户数据,进行全局最小化,以此对局部模型进行拟合,减

1. 甘肃同兴智能科技发展有限公司 甘肃兰州 730050

少模型丢失。同时，联邦学习还将对生成网络进行训练，并根据其隐含空间还原诱导分布，解决在训练中数据安全性的问题。

在完成多源异构数据整体特征空间分布的提取后，采用随机生成方法生成一系列噪声向量，通过产生网络形成迁移集合^[5]。在此基础上，分别基于局部学习模型和全局学习模型，对小规模迁移数据集进行融合，在考虑融合损失度的前提下，按照下述公式，将具有相似性特征的数据集进行整合。

$$R = L(y, \hat{y}) + \min V(D) \cdot \frac{1}{K} \sum_{k=1}^K P_k \quad (2)$$

式中： R 表示融合后的数据， L 表示损失度函数， $\hat{y}$ 表示样本均值， K 表示对抗训练次数， P 表示聚合函数， k 表示第 k 次训练。通过上述步骤，结合多次迭代，在整体模型的引导下，对多源异构市场营销数据进行融合，为后续数据一致性检验提供全面准确的数据依据。

2 基于差分隐私的异构数据校验矩阵构建

在多源异构电力市场数据一致性检验过程中，异构数据之间存在的差异会对检验结果的精度产生一定的影响。为了解决该问题，将差分隐私理论引入到数据一致性判定矩阵中，在保证数据保密性的前提下，实现高效的验证。

在保证数据时序特征的同时，满足差分隐私保护对数据干扰的要求^[6]。在此过程中，先建立一个初始的数据序列。将初始产生的数据行按照一定规则产生，并为以后的矩阵建立提供基本参照。

定义与累计数有关的运算方法，以产生新的数据行。针对特定的数据点，用公式计算为：

$$X(l) = \sum_{n=1}^l \alpha(n) \quad (3)$$

式中： X 表示生成的第 l 个数据列， α 表示原始数据点， n 表示数据编号。

在上述内容的基础上，将各观测点累计，得出各位置点的累计数值，从而反映出各时段的总体变化趋势。同时，为了更好地对数据进行平滑处理，构建邻接序列^[7]。通过计算当前累计值与先前累计值的加权和，得到连续序列中的各个元素。此过程用公式计算为：

$$X_{t+1}(l) = \frac{1}{2} \times [X_{t-1}(l) + X_t(l) + X_{t+1}(l)] \quad (4)$$

式中： $X_{t-1}(l)$ 、 $X_t(l)$ 、 $X_{t+1}(l)$ 表示前一序列累计值、当前序列累计值、下一序列累计值。通过上述公式，确保相邻序列集合可以更好地体现出数据的时间连续性及变化特性。

为了实现在一致性校验中的差分隐私处理，在构造连续序列时引入噪声，通过在每一个近邻序列上加入部分噪声，

比如拉普拉斯分布，保证数据的保密性^[8]。在加入噪声之后，按公式计算邻近序列值：

$$\dot{X}(l) = X_{t+1}(l) + \delta(l) \quad (5)$$

式中： $\dot{X}(l)$ 表示添加噪声后的相邻序列值， $\delta(l)$ 表示噪声序列。

以加噪后的邻接序列为样本，建立矩阵模型。将矩阵模型中的各参量作为待求解的参量，采用最小二乘法，将其转化、展开式，获得异质数据原序列的近似表达式。通过此种方式，实现差分隐私下异质数据校验矩阵的构造，为电力市场多源异质市场营销数据的一致性检验奠定坚实的理论基础。

3 数据一致性校验算法

在上述内容的基础上，利用构建的校验矩阵，结合数据之间的属性对应关系，进行数据的一致性校验。

在电力市场环境，各系统产生的数据是存在显著差异的，部分系统在数据库中所反映的参数包含“多维度”信息点参数，如用户基本信息、用电设备设定和用电信息等。数据的输入、设置、修改、删除等环节往往带有人为因素，以及异质系统缺少统一的维护手段，导致数据不一致、数据遗失等问题日益突出。

在不同的异质系统中，数据的定义方式有一定差异。以典型系统为例，其中一个系统所用的数据模型是属于逻辑模型的，其中数据根据对象、对象属性和对象间的关系对数据库中的信息进行描述，而逻辑数据则是通过对象类的概念表达，把用电量信息点划分成不同的对象。而另一种系统则定义了各种类型的数据表格，其之间的关系以分层的方式进行布置，用电数据的定义就是“设备表”或者“测点用电”信息表→“用电负荷”定义“预用电”的单向触发方式。

通过对两种异构系统的数据定义方式以及数据重要度的分析，确定数据衍生所需包含的参数。在对异构数据库特征进行分析的基础上，根据数据和数据之间的逻辑关系，将其分解、重组^[9]。根据校验矩阵，可以按照下述公式，进行数据集差异度的识别，以此精准校验其中的异常数据。此过程用公式表示为：

$$U = \frac{\dot{X}(l) |\eta_1 - \eta_2|}{\eta_1 + \eta_2} \quad (6)$$

式中： U 表示数据集差异度， η_1 、 η_2 表示两组数据的序列均值。

通过对各系统的相应值进行对比，如果差值>设置阈值，则判断对应数据未通过一致性校验。在此基础上，根据数据之间的属性对应关系，对不同类型的数据进行逐列比对，形成差异分析报告^[10]。以此种方式，实现数据一致性校验算法的设计。

4 对比实验

4.1 实验准备

为实现对该方法在应用中效果的测试，选择某大型电力营销服务中心作为本次实验的研究试点，根据实验的具体需求，按照表 1，进行测试终端测试机的配置与实验中相关参数的设置。

表 1 测试机配置与实验参数设置

序号	项目	参数
1	测试终端架构	基于中心化的系统架构
2	服务器	1 台聚合服务器
3	训练节点	20 个边缘训练节点
4	测试集：训练集	20:1
5	迭代次数（次）	600
6	批处理大小	32
7	学习率	0.01
8	优化器	Adam

在上述内容的基础上，在测试中心中的多功能标准表、耐压试验校验仪、绝缘电阻测试仪等，采集试验中的样本数据，在进行试验之前，要保证各项设备的性能良好，性能稳定，并做好相应的技术文件，以便进行试验。采样的部分样本数据如表 2 所示。

表 2 多源异构电力营销数据一致性校验中的样本数据

序号	数据类别	数据条数 / 条
1	用户基础数据	21 562
2	电能表读数数据	5 056 215
3	电费账单数据	52 015
4	业务申请数据	66 217
5	业务审批数据	561 025
6	服务工单数据	10 255
7	用电负荷数据	225 123
8	电费缴费记录数据	520 452

表 3 多源异构电力营销数据的一致性校验结果（单位：条）

数据源	异常数据 条数	本文方法		文献 [1] 方法		文献 [2] 方法	
		检验异常数据条数	召回数据条数	检验异常数据条数	召回数据条数	检验异常数据条数	召回数据条数
取值范围异常	125	125	125	123	123	110	110
数据类型错误	201	201	201	109	109	180	180
空值错误	23	23	23	23	23	18	18
符号错误	254	254	254	245	245	230	230
关系错误	195	195	195	185	185	200	200
其他错误	230	230	230	220	220	205	205

通过上述方式，完成本次实验的相关准备工作，在此基础上，引进文献 [1-2] 提出的方法，将其作为对照方法，进行多源异构电力营销数据一致性校验的相关测试。

4.2 电力营销数据一致性校验结果检验

在多源异构电力营销数据的一致性校验中，针对异常数据的回收和实际情况进行比较，检查异常数据条数。当二者之比趋于 1 时，说明检验算法能够更全面、更准确地发现异常数据，从而为电力市场营销数据质量的有效保证，为后续的数据分析和决策提供可靠的支持。

反之，如两者相差较大，说明对应的方法在进行多源异构电力营销数据的一致性校验中，未能实现对所有异常数据的识别，导致召回异常数据条数与实际异常数据条数存在较大偏差，影响电力营销的服务管理水平。以此为依据，对多源异构电力营销数据的一致性校验结果进行检验，如表 3 所示。

从表 3 所示的实验结果可以看出，应用本文设计的方法进行多源异构电力营销数据的一致性校验，该方法可以实现对所有异常数据的检验，并针对所有检验得到的异常数据进行召回，确保一致性检验结果的可靠性。

在此基础上，对文献 [1-2] 方法的检验结果进行分析可以看出，对应方法检验的多源异构电力营销异常数据与实际异常数据条数存在偏差，虽然文献 [1-2] 方法也可以实现对所有识别异常数据的召回，但由于检验的异常数据条数与实际不符，导致召回异常数据后的原数据质量也无法达到最优。

4.3 多源异构电力营销数据一致性校验中损失度检验

多源异构电力营销数据一致性校验中损失度检验，是指在进行数据一致性检验训练时，所造成损耗程度的度量。在多源异构数据融合和一致性过程中，数据的格式转换和融合算法等会造成数据信息的丢失或损坏。但随着训练轮次的增加，数据在训练中的损失度应呈现递减趋势，并在收敛后达

到一个相对较小的数值。如在训练过程中，损失度并未快速收敛或收敛后仍未能达到较低水平，说明在一致性校验过程中，数据存在丢失等异常现象。以此为依据，对三种方法的实验结果进行分析，如图 1 所示。

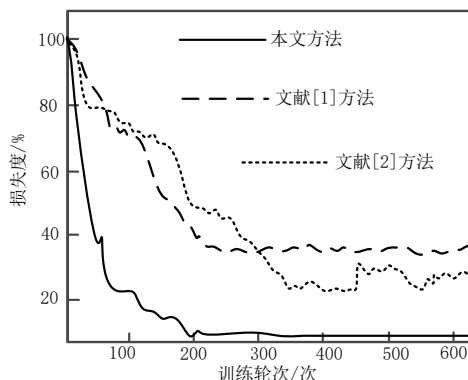


图 1 多源异构电力营销数据一致性校验中损失度检验

从图 1 所示的实验结果可以看出，设计的方法应用后，可以在多源异构电力营销数据一致性校验中，保证数据损失度快速收敛，并最终收敛至极低水平，避免数据的丢失等异常。

文献 [1-2] 方法虽然也在训练中，实现了数据的收敛，但无法达到本文方法应用后所达到的损失水平，因此可以认为在相同的测试条件下，应用文献 [1-2] 方法进行多源异构电力营销数据一致性校验，容易出现数据的丢失、损失等异常现象。

5 结语

本文通过基于联邦学习的多源异构电力营销数据融合、基于差分隐私的异构数据校验矩阵构建、基于属性对应关系的数据一致性校验算法，开展了此次研究。在后续的工作中，将结合电力行业营销工作的具体需求，尝试将该方法进行推广应用，为提高电力营销保障等提供全面的技术支持。

参考文献：

- [1] 孔祥磊,汪芳琴,钟选明,等.具有方向约束的多智能体系统的反一致性研究[J].空间控制技术与应用,2023,49(1):74-81.
- [2] 卫泽晨,李立新,刘金波,等.基于改进默克尔树与区块链的电网调度自动控制软件版本一致性管控方法[J].电网技术,2024,48(3):1273-1280.
- [3] 许亮业,石连星,单蓉胜.一种纠删码条带化数据的一致性检查方法[J].上海交通大学学报,2024,58(4):579-584.
- [4] 胡伟洋,孙英云,赵鹏飞,等.基于日电量平衡与日内校验的长期随机运行模拟算法[J].电网技术,2023,47(10):

4263-4271.

- [5] 桑建斌,李玉平,丁俊健,等.基于递归型定频采样变数据窗离散傅氏算法的大型旋转电机启机保护[J].电力系统保护与控制,2023,51(6):137-144.
- [6] 黄梓欣,汪涛,徐昂,等.基于临界超标样本扩充的数据驱动短路电流超标精准校验方法[J].电网技术,2024,48(6):2564-2573.
- [7] 李丹,秦世耀,李少林,等.基于混沌粒子群的双馈风电机组 LVRT 实测建模及暂态参数辨识[J].中国电力,2024,57(8):75-84.
- [8] 黄闽南,王一山,范佳铭,等.基于方差阈值的分布式光纤振动定位算法研究[J].传感技术学报,2024,37(12):2113-2118.
- [9] 姚钦才,向文国,陈时熠,等.基于 ICEEMDAN-KPCA-ICPA-LSTM 的光伏发电功率预测[J].动力工程学报,2025,45(3):374-382.
- [10] 董琴,杨涛.基于 RBF 神经网络的关联数据一致性挖掘仿真[J].计算机仿真,2023,40(7):457-461.

【作者简介】

戴媛媛（1980—），女，江苏兴化人，硕士研究生，中级工程师，研究方向：财务管理。

司健宏（1990—），男，甘肃定西人，本科，中级工程师，研究方向：电力营销。

梁锐凯（1992—），男，甘肃兰州人，本科，助理工程师，研究方向：计算机科学与技术。

（收稿日期：2025-10-10 修回日期：2026-05-12）