DOI:10.20033/j.1003-7241.(2025)06-0076-04

基于人工智能技术的电力信息运维数据整合平台

郭 蕊，李奕霏，高育栋
（国网甘肃省电力公司数字化事业部，甘肃 兰州 730050）

摘 要：为保证电力信息运维数据整合效果，设计一种基于人工智能技术的电力信息运维数据整合平台。采用基于密度峰值的聚类算法找出数据中的异常值，采用人工智能技术中的机器学习算法清洗数据，以消除数据中的异常点，避免异常值对整合过程产生影响。最终，经数据编码、抽象及重组后，运用深度学习算法集成电力运维数据。实验结果证明，该方法精度高、效率高、覆盖广，具有一定的实用价值。

关键词：密度峰值聚类法；自回归模型；数据清洗；深度学习算法；数据编码；抽象矩阵
中图分类号：TP181 文献标识码：A 文章编号：1003-7241(2025)06-0076-04

Power Information Operation and Maintenance Data Integration Platform Based on Artificial Intelligence Technology

GUO Rui, LI Yi-fei, GAO Yu-dong
(Digitalization Department of State Grid Gansu Electric Power Company, Lanzhou 730050, China)

Abstract: In order to ensure the integration effect of power information operation and maintenance data, this paper designs a power information operation and maintenance data integration platform based on artificial intelligence technology. The clustering algorithm based on the density peak is used to find the outliers in the data. The machine learning algorithm in artificial intelligence technology is used to clean the data to eliminate the outliers in the data and avoid the impact of outliers on the integration process. Finally, after data encoding, abstraction, and recombination, deep learning algorithms are used to integrate power operation and maintenance data. The experimental results demonstrate that this method has high accuracy, high efficiency, and wide coverage, and has certain practical value.

Keywords: density peak clustering method; autoregressive model; data cleaning; deep learning algorithm; data encoding; abstract matrix

0 引言

随着现代科学技术的发展与大数据时代的来临，数据整合技术逐渐成为热门领域，被广泛应用于互联网能源、电力分析、医疗研究[1]等领域。为了创造安全高效的电力运营环境，我国工作人员对电力信息运维数据的整合方法提出了更高的要求，不仅创建出集数据存储与标记为一体的数据整合库，还拓展出元数据与异构数据的总整合框架。国外在数据整合方面已完成领域之间的过渡研究，将数据整合技术应用于医疗、计算机、交通规划等方面。

杨舒[2]等人创建一种基于微服务的电力信息运维数据整合系统。在系统中构建云服务架构，保证数据之间时刻保持均衡状态，最后采用轻量级数据整合算法，通过数据间的可视化分布完成电力信息运维数据的整合。但该方法没有剔除数据中的异常值，导致方法的整合精度低。方匡南[3]等人首先采用近邻法构建电力信息运营数据的网络结构，同时采用惩罚因子识别数据中的异常数据，其次通过最短路径代价法剔除异常数据，最后在参数估计与分类预测的模拟设置下完成电力信息运维数据的整合。但该方法没有对数据做聚类处理，导致方法的整合效率低。龙北平[4]等人设计了一种基于权重相似性模型的不动产数据整合方法。首先，通过欧氏距离和相似度系数建立相似离度，根据确定关联目标实现地块的落宗匹配，完成数据的整合。但该方法未对数据进行编码处理，导致覆盖度较低。

因此，本文设计了基于 AI 的电力信息运维数据整合平台，以解决现有问题。

1 电力信息运维数据预处理

1.1 数据聚类

为了避免电力信息运维数据中的异常值对整合过程产生影响，采用基于密度峰值的聚类算法找出数据中心的异常值。具体步骤如下：

*基金项目：中国南方电网有限责任公司科技项目（000000KK52150012）
收稿日期：2023-11-30

(1) 基于密度峰值的聚类算法首先计算电力信息运维数据的局部密度,公式如下所示:

$$r=\sum_{q} z\left(D_{pq}-D'\right) \tag{1}$$

式中,r为数据的局部密度值;z为电力信息运维数据点;D为欧式距离;D'为截断距离;p、q均为数据集合中的任意点。

(2) 在局部密度中引入阈值密度函数[5],保证数据点z与周围数据点之间的距离均小于截断距离D',阈值密度函数α公式如下所示:

$$\alpha\left(z\right)=\begin{cases}1, & z\le 0\\ 0, & z>0\end{cases} \tag{2}$$

(3) 计算数据点z与局部密度值大于z的点之间的距离,得到电力信息运维数据之间的总距离,公式如下所示:

$$\beta=\begin{cases}\min D_{pq}, & r_p<r_q\\ \max D_{pq}, & r_p\ge r_q\end{cases} \tag{3}$$

式中,β为电力信息运维数据之间的总距离。

(4) 当电力信息运维数据中存在到类簇距离大于β、密度值小于α的点时,此时聚类计算结束,判定该点为异常点。

1.2 数据清洗

针对聚类得到的电力信息运维数据异常点,采用人工智能技术中的机器学习算法清洗数据,消除数据中的异常值。具体步骤如下:

(1) 将电力信息运维数据视为具有平稳序列的数据组z_u,然后输入到自组织神经网络模型中实现数据的拟合,公式如下所示:

$$\begin{cases}z_u=\chi z_{u-1}+\delta\\ u=1,2,\cdots\end{cases} \tag{4}$$

式中,χ为自组织神经网络模型的回归系数;δ为数据中的序列表达;u为电力信息运维数据的整体数据量。

(2) 自组织神经网络模型中的回归系数与序列表达的计算公式分别如下所示:

$$\begin{cases}\chi=\dfrac{\varepsilon_2}{\varepsilon_1}\\ \delta=\varepsilon_2-\varepsilon_1\chi=\dfrac{\sum\limits_{u=1}^{u} z_u^2}{u}-\dfrac{u\cdot\left(\sum\limits_{u=1}^{u-1} z_{u+1}\right)^2}{\sum\limits_{u=1}^{u} z_u^2}\end{cases} \tag{5}$$

式中,ε_1、ε_2分别为电力信息运维数据序列的方差与方差系数。

(3) 在自组织神经网络模型中,判断神经元对应的权值与数据的相似性,然后根据先验知识优化权值[6],在保证其他神经元不变的前提下输出权值,公式如下所示:

$$\begin{cases}w_l\left(u+1\right)=w_l+\phi\left[z_u-w_l\left(u\right)\right]\\ w_l\left(u+1\right)=w_l\end{cases} \tag{6}$$

式中,w_l为神经元l的权值;ϕ为学习率。

(4) 经过自组织神经网络模型输出的权值即为电力信息运维数据不同类别的中心数据值,将与数据点z最接近的权值作为量化值的索引,得到完全均匀量化的电力信息运维数据,此时异常值被消除,完成数据的清洗。公式如下所示:

$$\begin{cases}O_u\in\left\{w_1,w_2,\cdots,w_m\right\}\\ o_u=O_{h(z)}\\ O_{h(z)}\in\left\{1,2,\cdots,m\right\}\end{cases} \tag{7}$$

式中,O_u为量化值的集合;w为权值;O_u为量化值;$O_{h(z)}$为量化值对应的索引;w_m为的是第m个权值。

2 数据整合平台

采用深度学习算法整合电力运维数据,步骤如下:

(1) 基于深度学习的数据整合平台,主要包含资源层、核心层与服务层,结构如图1所示。其中资源层负责数据编码与数据访问;核心层负责集成、优化、抽象表达与重组数据;服务层负责提供可扩展标记语言并生成对应的服务。

(2) 采用不同的编码准则[7-8]对电力信息运维数据编码,保证运维数据与编码之间一一对应,公式如下所示:

$$U=\frac{e}{\sum\limits_{i=1} v\cdot k} \tag{8}$$

式中,U为编码后的电力信息运维数据;e为数据的不同类别表示;v为编码参数;i为编码次数;k为编码顺序。

(3) 在公式(8)的基础上改进编码能力[9],提高电力信息运维数据的精度与归一性,为数据整合打下基础,公式如下所示:

$$|f|=\int\left(\frac{s}{e}\right)\cdot U \tag{9}$$

式中,$|f|$为编码准则;s为电力信息运维数据中的规范数据。

(4) 设电力信息运维数据集合由M个数据构成,数据集表达为$L=\{L_0,L_1,\cdots,L_j,\cdots,L_M\}$,则该数据集可以用$N$维矢量构成的集合$g$表示,公式如下所示[10]:

$$g=L\left(g_{0,j},g_{1,1+j},\cdots g_{N-1,j+M}\right)^V \tag{10}$$

式中,$g_{0,j}$为数据集中第j个节点存储的0个运维数据;$g_{1,1+j}$为第$1+j$个节点存储的1维数据;$g_{N-1,j+M}$为第$j+M$个节点存储的$N-1$个运维数据;V为矢量集合。

(5) 用抽象矩阵表达原始运维数据,将节点数据抽象

为集合g^1，挖掘数据的相关性，公式如下所示[11-12]：

$$\begin{cases} g^1=\begin{bmatrix} g_{0,0} & \cdots & g_{0,M-1} \\ \vdots & \ddots & \vdots \\ g_{N-1,0} & \cdots & g_{N-1,M-1} \end{bmatrix} \\ g_{a,j}^{1,E}=\sum_b E_{b-2a}g_{b,j},0\le a\le\dfrac{M}{2} \\ g_{a,j}^{1,F}=\sum_b F_{b-2a}g_{b,j},0\le a\le\dfrac{M}{2} \end{cases} \tag{11}$$

式中，$g_{a,j}^{1,E}$、$g_{a,j}^{1,F}$均为节点j处的离线数据；a为整合后的电力信息运维数据；b为原始数据；E、F均为相关系数。

(6) 通过数据重组得到运维数据排列矩阵，在保证原有数据结构不变的同时，实现多个数据的转换，完成电力信息运维数据的整合，公式如下所示：

$$g_1=\begin{bmatrix} g_{0,0}^{1,E} & g_{0,1}^{1,E} & \cdots & g_{0,M-1}^{1,E} \\ \vdots & \vdots & \vdots & \vdots \\ g_{\frac{N}{2},0}^{1,E} & g_{\frac{N}{2},1}^{1,F} & \ddots & g_{\frac{N}{2},M-1}^{1,E} \\ g_{\frac{N}{2},0}^{1,F} & \cdots & \cdots & g_{\frac{N}{2},M-1}^{1,F} \end{bmatrix} \tag{12}$$

式中，g_1代表数据排列矩阵。

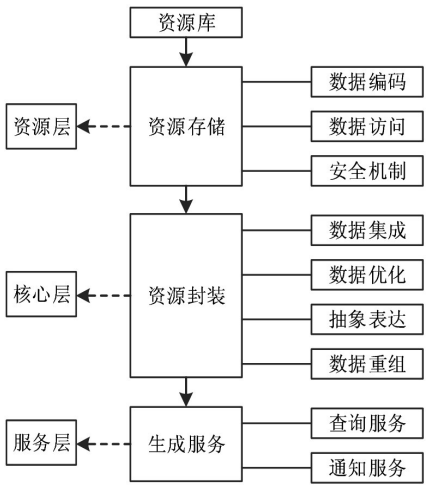


图1 电力信息运维数据整合平台结构

3 实验与分析

为验证AI运维数据整合平台的有效性，以整合精度、效率和数据覆盖度为评价指标，与文献[2]、文献[3]、文献[4]方法进行对比测试。

(1) 整合精度

AUC是评价整合精度的重要指标，值越大，整合精度越高；反之越低。

绘制ROC曲线，根据对比AUC值以评估各方法(所提方法、文献[2]、文献[3])的整合精度。其中假阳性率作为横坐标、真阳性率作为纵坐标。公式如下所示：

$$\begin{cases} FPR=\dfrac{TP}{(FN+TP)} \\ TPR=\dfrac{FP}{(FP+TN)} \end{cases} \tag{13}$$

式中，FPR为假阳性率、TPR为真阳性率；TP为整合过程中，将负向数据整合成负向数据的数据量；FN为将负向数据整合成正向数据的数据量；FP为将正向数据整合成负向数据的数据量；TN为将正向数据整合成正向的数据量。

测试结果用图2表示。

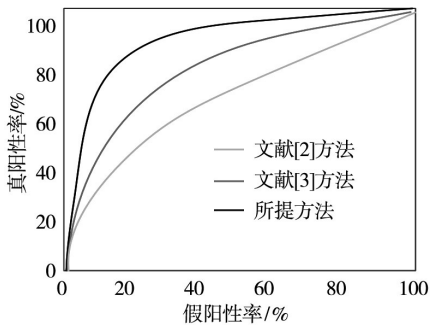


图2 不同方法的ROC曲线

分析图2可知，所提方法的AUC面积大于文献[2]方法与文献[3]方法的AUC面积，说明针对电力信息运维数据的整合，所提方法的整合精度高于文献[2]方法与文献[3]方法的整合精度。

所提方法在电力信息运维数据的整合过程中，首先采用密度峰值聚类算法找出数据中的异常值，并采用机器学习算法消除数据中的异常值，进而在整合过程中不受异常数据的影响，提高了方法的整合精度。

(2) 数据整合效率

对比所提方法、文献[2]、文献[4]整合电力运维数据的效率，在相同的数据量中，整合时间越短效率越高。其计算公式如下所示：

$$t_x=\eta\times\kappa+\mu_x\times\kappa \tag{14}$$

式中，t_x为第x次整合需要的时间；η为完整周期内的数据出块个数；κ为数据出块时间；μ_x为第x次整合时的数据量。

表1 不同方法的整合时间

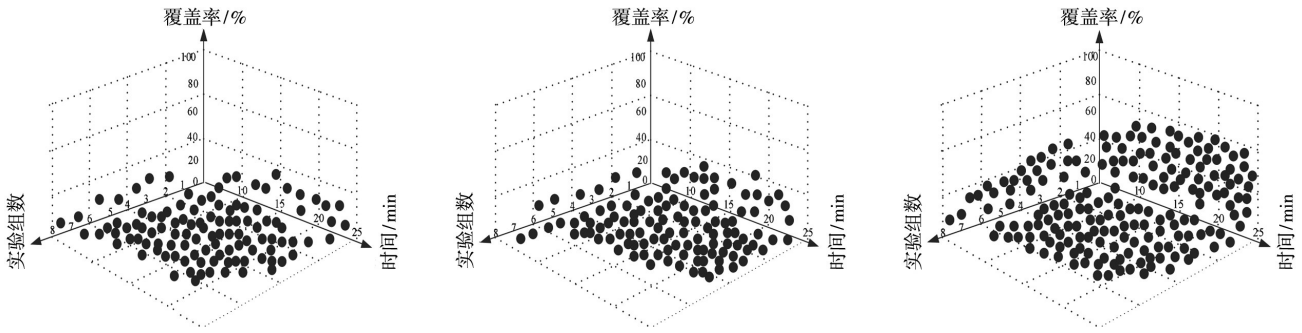
数据数量/组	整合时间/s			数据数量/组	整合时间/s		
	所提方法	文献[2]方法	文献[4]方法		所提方法	文献[2]方法	文献[4]方法
10	33	43	60	90	52	76	94
20	36	47	63	100	55	79	99
30	37	50	67	110	59	83	103
40	40	53	71	120	62	86	109
50	42	58	74	130	64	89	114
60	44	62	78	140	68	95	121
70	47	67	82	150	71	102	125
80	51	71	87				

其整合的效率测试结果如表1所示。

分析表1可知，整合时间与电力信息运维数据数量之间成正比，随着数据数量的不断增加，所提方法、文献[2]方法和文献[4]方法所用的整合时间不断增加。在相同电力信息运维数量下，所提方法的整合时间均低于文献[2]和文献[4]，表明针对电力信息运维数据的整合，所提方法的整合效率高于文献[2]方法与文献[4]方法的整合效率。

(3) 数据覆盖度

数据覆盖度是指整合后的数据占整体电力信息运维数据的分区情况。其值越大，表明方法整合后的数据越全面；其值越小，表明方法整合后的数据越单一。将所提方法、文献[3]方法与文献[4]方法的数据覆盖度测试结果绘制成图3，以便分析。



(a) 文献[3]方法　(b) 文献[4]方法　(c) 所提方法

图3 不同方法的数据覆盖度

由图3可知，所提方法整合电力运维数据的覆盖度高于文献[3]和文献[4]，表明其数据更全面。

4 结束语

为提升电力信息运维数据整合效果，本文设计一种基于人工智能技术的电力信息运维数据整合平台。该平台首先采用密度峰值聚类算法找出数据中的异常值，其次采用人工智能技术中的机器学习算法清洗异常数据，最后通过深度学习算法实现电力信息运维数据的整合。在提高了方法整合精度与整合效率的同时，也一定程度上提高了方法的数据覆盖度，为数据整合效果的提升奠定了一定基础。

参考文献：

[1] 钟雅婷，林艳梅，陈定甲，彭昱忠，曾远鹏.多组学数据整合分析和应用研究综述[J].计算机工程与应用，2021，57(23)：1-17.

[2] 杨舒，苏放.基于微服务的分布式数据安全整合应用系统[J].计算机工程与应用，2021，57(18)：238-247.

[3] 方匡南，张晴雯，林洪伟.考虑数据源网络结构的高维数据整合分析与子群识别研究[J].统计研究，2022，39(7)：125-136.

[4] 龙北平，吴家杰，郑勤华，等.一种基于权重相似性模型的不动产数据整合方法[J].测绘通报，2021(6)：122-126.

[5] 贾象阳，黄先锋，牛文渊，等.顾及噪声密度函数差异的自适应回归算法及其在人工标靶提取中的应用[J].测绘学报，2021，50(2)：226-234.

[6] 叶霞，刘睿珩，许飞翔，等.基于SOA和本体的指挥信息系统数据集成研究[J].火力与指挥控制，2022，47(3)：136-143.

[7] 李国权，周湘云，林金朝，等.基于迫整检测的全分集多用户空时分组编码设计准则[J].通信学报，2019，40(3)：139-146.

[8] 王玥，苏寒松，刘高华.改进的编码-解码时序动作检测算法[J].激光与光电子学进展，2021，58(20)：463-469.

[9] 王磊，张晓东，樊博，等.电网企业营销数据综合治理系统设计[J].自动化技术与应用，2023，42(9)：148-151.

[10] 罗长银，陈学斌，刘洋，等.基于联邦集成算法对多源数据安全性的研究[J].计算机工程与科学，2021，43(8)：1387-1397.

[11] 汪斐，黄丽琼.基于大数据的英语学习投入量自动计算模型[J].自动化技术与应用，2023，42(2)：158-161.

[12] 方匡南，张晴雯，林洪伟.考虑数据源网络结构的高维数据整合分析与子群识别研究[J].统计研究，2022，39(7)：125-136.

作者简介：郭蕊(1995—)，女，硕士，助理工程师，研究方向：电网数字化建设运营。