**1台区线损异常数据库建立**

现场核查定位线损异常台区的影响原因及用户明细，获得真实标签数据。根据线损正常台区和线损异常台区数据，建立台区线损异常数据库，用于有监督机器学习的训练数据集。

**2电能表状态评价系统**

电能表状态评价系统根据“能量守恒”原理对低压台区内电能表的计量状态进行评价，本质上分析台区线损与电能表电量的关系（以相关性角度挖掘异常、皮尔逊相关系数）。根据模型输出的台区线损率，电能表计算误差值，置信度指标，输出电能表排查清单。

**3线损异常特征聚类**

通过提取特征形成对台区及核查用户的多维度描述，用以辅助类似台区及用户的分析，按照无监督的聚类思路进行相似案例的判别分析。由于用户的用电负荷数据维度多、数据量大，为了更高效、准确的对用户的用电行为进行聚类分析，需要对原始用电数据进行降维，即特征提取。

对提取的特征采取DBSCAN算法进行聚类分析，并制定一定的规则，对台区及用户进行预分类，进行初筛，提高用电行为分析模型的效率。

DBSCAN算法具有足够密度的区域划分为簇，并在具有噪声的空间数据库中发现任意形状的簇，它将簇定义为密度相连的点的最大集合。可以对任意形状的稠密数据集进行聚类；可以在聚类的同时发现异常点，对数据集中的异常点不敏感；聚类结果没有偏倚。

该聚类算法根据密度进行划分，可以针对不同的台区用电行为进行聚类分析，受异常噪声数据影响低，具有很强的鲁棒性。

通过特征工程提取以下用电特征，用以用电行为分析。

表1 台区日用电特征

|  |  |
| --- | --- |
| 特征 | 特征 |
| 日用电负荷 | 峰谷差率 |
| 日平均负荷 | 尖期负载率 |
| 日最大负荷 | 峰期负载率 |
| 日最小负荷 | 平期负载率 |
| 负荷率 | 谷期负载率 |
| 峰谷差 | 线损率 |

表2 用户用电特征

|  |  |
| --- | --- |
| 特征 | 特征 |
| 用电均值 | 电能表安装日期 |
| 用电标准差 | 用电性质 |
| 最大有功负荷 | 周电量曲线 |
| 无功负荷 | 月电量曲线 |
| 低负荷占比 |  |

针对用户的节假日属性，提取以下特征。

表3 节假日特征

|  |  |
| --- | --- |
| 特征 | 特征 |
| 工作日日平均负荷 | 工作日负载率 |
| 节假日日平均负荷 | 节假日负载率 |

**4大数据自学习系统**

部署有监督和无监督机器学习模型，对于闭环的大量真实标签数据，通过人工提取台区和用户的用电特征或者特征工程，根据标签信息训练有监督模型，对于待分析台区或用户，通过模型即可得出是否存在异常以及何种异常。同时，训练无监督模型，对于待分析台区或用户进行类别初判或者预分析，辅助有监督模型的判断。随着有效闭环工单的累积，训练样本量增多，涵盖的场景越来越广泛，通过对模型实时训练更新，使得模型具有自适应性及完备性。

利用大数据的优势，将数据挖掘技术和机器学习相结合，其中神经网络具有强大建模能力和表征能力，能够挖掘台区线损异常特征的内在联系，解决人工排查、常规逻辑判断的局限性，实现台区线损异常分类。

下图为大数据自学习系统的简要过程：

