

[大数据]

基于大数据的电力用户信用评估体系构建

文/顾安朋
中国南方电网有限责任公司

伴随着人们对电力服务需求的多样化以及客户对体验式及自助式服务需求日益迫切，供电营业厅通过部署多功能的营业厅智能设备满足客户服务办理需求。随着自助服务增多，如何防范与应对违约风险成为大数据时代供电营业厅的重要研究课题。基于此，以大数据为背景，依托智能算法、数据分析与处理等先进技术创建了一套电力用户信用评价体系，通过分析电力大数据，预测未来可能出现的违约风险并针对客户既有行为进行信用评级。验证结果表明，这套电力用户信用评价体系支持用户用电特征的精准识别，同时为供电营业厅规避违约风险提供了数据支持。最后，对于电力企业应对用户违约风险提出了几点建议，与信用评估体系相结合，希望可进一步提升风险防范与应对能力，减少违约行为。

伴随着社会信用体系广泛推广，电力系统也逐步建立起对应的信用体系，智能算法与模型的应用，在一定程度上约束用户用电行为，减少违约。本文秉承“表扬诚信，惩罚失信”的原则，综合运用多种用户用电特征识别方法，如K-means聚类方法、回归模型等，构建相应的电力用户信用评价体系，以期对引导电力用户积极缴费，减少违约等有所帮助。

研究背景

在新经济发展趋势下，电力行业着手推动的信用体系建设转型，是建立品牌形象，实现创新发展的必然选择。随着电力市场持续深化改革，社会信用体系建设日益完善，电力行业有必要基于大数据构建电力用户信用体系，以应对新挑战。

广东省积极响应国家号召，按照世界银行和国家发展和改革委员会营商环境评价标准，主动对标世界一流供电企业，深入查找差距和短板，不断开创广

州办电加速度，优化供电服务新体验。广东供电局为满足客户便捷化业务办理需求，全面部署多功能营业厅智能设备，开发了征信办电、设备监控和多版本软件功能。随着项目持续推进，综合运用大数据、智能算法等技术构建合理有效的电力用户信用评价体系显得尤为重要。

电力用户信用评价体系

1. 电力用户用电信用评估指标体系

当供电企业与用户建立起交易关系后，一旦双方中任意一方作出任何违约行为，都会引发严重的不良后果，对此，提出了审核与制约交易双方行为的解决方案。因供电企业生产具有一定的稳定性，所以，通常更加关注用户也就是购电方的交易行为，也就是说，需要面向全体用户构建相应的信用评估体系，落实信用评估行为，划分用户信用等级、风险等级，针对评估结果制定个性化的电力交易方案。电力用户用电信用评估指标见表1，共有5

个一级指标，23个二级指标。

表1 电力用户用电信用评估指标体系

一级指标	二级指标	居民	低压非居	高压非居
商业信用	电量增长率	✓	✓	✓
	区域容量占比	✓	✓	✓
	电费贡献率	✓	✓	✓
安全信用	实名制认证信息	✓	✓	✓
	安全事故次数			✓
	安全事故恶劣程度			✓
	预防性电器试验超周期			✓
	未配置制证高压电工			✓
行为法律信用	违章用电次数	✓	✓	✓
	窃电次数	✓	✓	✓
	通知整改次数	✓		
	窃电度数	✓		
	违约用电违约金拖欠金额	✓		
	电费支付方式（代扣）	✓		
	电费支付方式（托收）		✓	✓
	违章用电平均恶劣程度（超容）		✓	✓
	违章用电平均恶劣程度（私转供）		✓	✓
经济法律信用	通知整改次数（高价低接）		✓	✓
	欠费停电次数	✓	✓	✓
	欠费次数	✓	✓	✓
	欠费金额	✓	✓	✓
	欠费时长		✓	✓
合作信用	欠费停电总时长		✓	✓
	客户保修客观性	✓	✓	✓
	调度合作记录		✓	✓

2. 电力用户信用评估指标体系优化设计

在上述电力用户信用评估指标体系的基础上，利用信用评分建模历史经验，结合用电用户基本特征，从守信能力的角度进一步优化电力用户信用评估指标体系，以强化指标的实用性。

基于守信能力视角优化电力用户信用评估指标体系，具体操作时需利用大数据技术评估用电价值以及电网互动价值。

关于用电价值评估：其指标体系如图1所示。考虑重点是行业水准和电量电费，强调用户用电价值的客观评价。

关于电网互动价值评估：其指标体系如图2所示。其指标包括源网荷互动、需求侧互动以及电能替代互动。电网互动价值评估结果，可为供电企业与用户之间的良性互动提供参考。

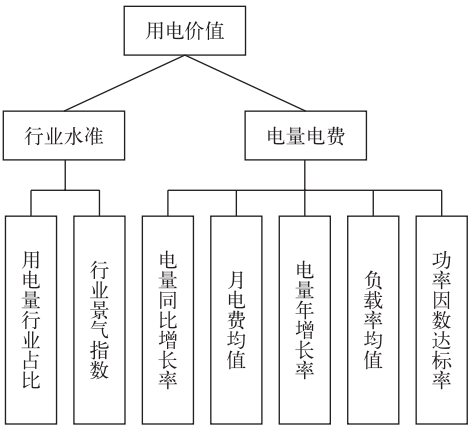


图1 用电价值指标体系

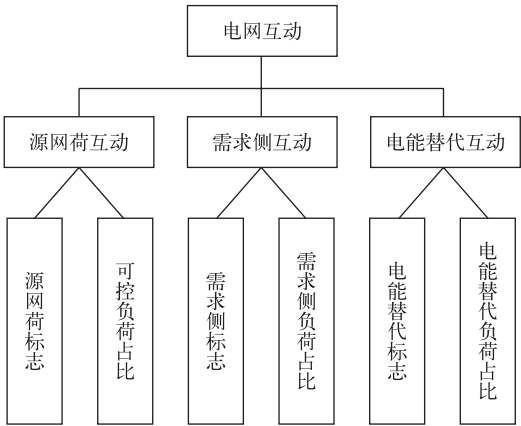


图2 电网互动价值指标体系

为保障电力用户信用评价体系的有效性，综合运用处理算法，如神经网络、灰度关联分析法和遗传算法等，其流程图如图3所示。

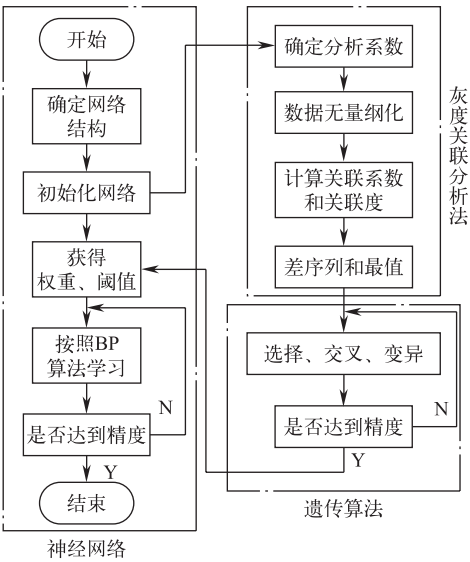


图3 信用评估体系中综合算法处理流程图

基于电力大数据的电力用户信用评价模型关键技术

1. 用户特征识别方法

构建电力用户信用评估模型前，应基于电力大数据，运用数据分析与处理技术多维度提取用户用电特征，分类管理，深入分析，观察不同时段用户用电特征。

(1) K-means聚类方法

针对用户用电特征提取需求，引入了K-means聚类方法，通过优化最小化目标函数获取局部最佳聚类效果，从而精准获取用户用电特征。

在采用K-means聚类方法时，为保障聚类效果，应在反复挑选不同原始凝聚点的基础上迭代计算。

假设：原始数据集中存放了 m 个样本，已知样本特征属性为 Z 且 $Z > 0$ ，聚类运算时的推算解析式为 $F = \sum_{i=1}^z F_i = \sum_{i=1}^z \sum_{j=1}^m \alpha_{ij} \|A_j - B_i\|^2$ ，其中， F 、 A_j 、 B_i 、 α_{ij} 分别表示聚类中心距离平方和的目标函数、 j 个用电数据特征属性、 i 个特征聚类中心点和权重指数。

(2) 回归模型

假设：电力用户历史信用评分数据集为 (x_i, y_i) ，其中， $i \in N$ ， x_i 是指描述电力用户信息的列表， x_i 是指用于识别用户用电特征的标签，在提取用户用电特征时，可构建 $y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$ 的回归模型，其中， β_0 、 β_j 、 x_{ij} 分别表示截距系数、回归系数和列表中的数据。

在此基础上，可采用Lasso模型简化回归模型。

假设：每个观测值是相对独立的个体，彼此之间互不相关，且变量标准化，此时，由 $\hat{\beta} = \arg \min \left\{ \sum_{i=1}^N (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$ 可以看出 β_j 中系数的缩减与 λ 有关，二者之间的关系是 λ 随

着系数逐渐减小并接近0而增大。采用Lasso模型时，支持系数缩减与特征选取同步进行。

(3) 数据预处理

无论是选用哪一种用户用电特征提取方法，考虑到外界噪声扰动，后续均需要进行数据预处理，以提高数据质量。

第一步：数据清洗。通过观察两个记录同时处于某个属性的关联性，判断二者相似度，经过累加计算清理数据中的偏离集群点，以提高数据的完整性。

假设：用电数据记录 R 中某个数据 R_i 与数据记录数量 l 相对应，其关联表达式为 $M = A(R_{ip}, R_{jp}) = \{0, 1\}$ 。与1相对比，若 $M=1$ ，则表明两个记录均属于 P 属性，此时，两条数据高度相似，关联性强。反之，表明数据之间不存在相似关联性。确定某属性值下两个用电记录的相似性，可运用公式 $Q_{i,j,p} = MW_p$ （用电记录内 P 属性权重）。运用公式 $Z_{i,j} = \sum_{p=1}^n Q_{i,j,p}$ 累加求和各用电记录属性相似度。

第二步：数据集成。数据集成的前提是提高数据质量，消除冗余数据，减少干扰。引入元数据策略，降低数据集内错误识别概率。

第三步：数据规约。为有效规约大规模数据集，对数据进行规约处理，具体操作为在充分考虑数据随机变量的基础上，映射初始数据，使其被放置在小范围空间内；采用数据压缩评判，以明确重构初始数据的可行性。

第四步：数据变换。为应对数据噪声，进行数据变换，通过更改属性、分割数据等方式，调整聚类分析难度。又或是采用自适应孤立森林噪声方法（AIFNM），先计算异常程度分数辅助噪声检测，后经过训练集中，增强噪声数据适应性，减少拟合现象。其执行过程如图4所示。

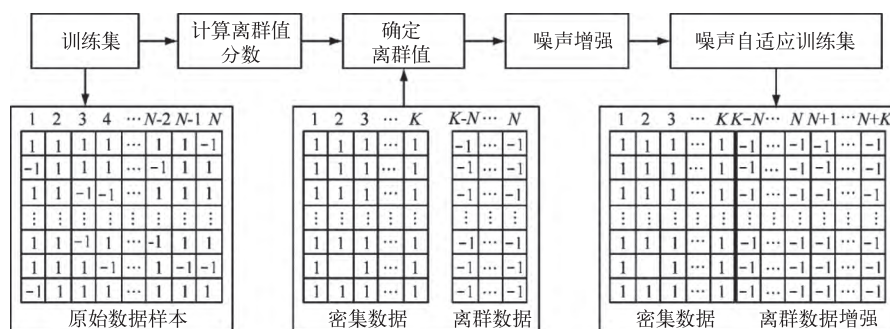


图4 AIFNM执行过程流程图

2. 灰度关联分析算法

在电力用户信用评估模型中，灰度关联分析算法负责多因素数据处理、数据关联度大小的计算以及次序分析。搭配遗传算法、神经网络算法，可实现数据的深入分析，明确不同影响因素的设计权重以及适应度，从而形成科学合理的信用等级。

采用灰度关联分析法时，先从多个影响因素中选取变量因素建立数据比较矩阵。在矩阵中包含多家企业的多个影响因素。采用无量纲化优化比较矩阵，以降低数值量的影响。然后，采用的差序列进一步处理数据，获取数据差值，形成差值矩阵。分析差值矩阵，明确最大和最小的各元素差值，经计算后获取数据关联系数，其中，采用关联系数计算公式为 $\xi_{01}(k) = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{oi}(k) + \rho\Delta_{\max}}$ ，其中， $\Delta_{\min}$ 、 $\Delta_{\max}$ 分别代表最小差值和最大差值， ρ （分辨系数） $\in [0,1]$ 。通过平权处理后，可知晓权重分配，采用层次分析法后，可明确灰度关联性算法中综合权重分配情况。

3. 进化网络模型算法

进化网络模型算法以数据训练验证研究结果误差，也就是说，采用进化网络模型算法可最大限度地保障用户信用评估结果的精度和准确性。在电力用户评估体系中采用进化网络模型算法，可获得关于各影响因素的学习误差函数 $E = \frac{1}{2} \sum_{k=1}^n \sum_{i=1}^p (y_{ki} - o_{ki})^2$ ，其中， y_{ki} 、 o_{ki} 分别表示算法中在理想情况下本研究所输出的结果、在算法中实际所输出的结果； n 、 p 分别表示本研究中需要输入的自变量影响因素的数据量、通过该算法需要输出的数据的量。基于学习误差函数，也可获得适应度函数 $\text{fitness} \frac{1}{E}$ 。结合权重向量等儿科进一步计算出样本基因变异概率。

仿真验证

在构建完电力用户信用评估体系后，利用几家不同电力采购企业的相关信息进行了验证。其主要验证内容是商业信用安全、信用行为、法律信用、经济法律信用和合作信用。经计算后，获得了灰度关联算法、综合算法等多个评估算法的正确率，分别是85%、96%。结果表明，基于混合算法模型设计的电力用户信用评估模型具有一定的可行性，可满足电力企业规避用户违约风险的实际需求。

用电用户信用等级体系

根据算法模型提供的数据分析与处理结果，结合指标体系，应科学划分用电用户信用等级，执行打分制，以更加直观地了解不同用户的信用等级以及变化情况，详细内容见表2。

表2 用电用户信用等级评定

信用等级		分值	信用状况
A	3A	100	信用极好
	2A	[90,100)	信用较好
	A	[80,90)	信用良好
B	3B	[70,80)	信用一般
	2B	[60,70)	信用欠佳
	B	[50,60)	信用较差
C	3C	[40,50)	信用很差
	2C	[30,40)	信用极差
	C	30以下	毫无信用

用户用电违约风险防控策略

电力企业在构建和推广电力用户信用评估体系的基础上，应根据用电用户信用等级制定对应的风险防控方案，以充分发挥信用体系作用，有效规避风险。

第一，针对守信用户：供电方可给予优惠政策，提供技术服务免费等优惠服务。完善奖励机制，酌情奖励。对于信用一般的用户，供电方应了解其经济能力，评估将其转化为优质守信用户的可行性，加大守信管理力度。第二，对于失信用户：供电方应制定差异化管理方案，动态跟踪用电情况，密切关注缴费进度等，组织其参加守信培训，提高守信意识。第三，依托物联网、人工智能等先进技术，搭建远程电量电费管控平台，注重监管40分及以下用户的用电缴费行为，采取强制措施，引起失信用户对用电缴费的重视，防范风险。

结束语

电力用户信用评估体系是供电方规避经营风险，加强用户管理的有效手段。因此，供电方应建立健全科学合理的评估指标体系，综合利用多种算法模型搭建电力用户信用评估模型，精准分析与处理数据，为供电方划分用户等级提供依据。根据用户信用评级采用对应的防范应对策略，以有效规避风险。EA

（收稿日期：2024-07-30）